

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ  
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА»

**Институт информационных технологий и технологического образования**  
**Кафедра информационных технологий и электронного обучения**

Основная профессиональная образовательная программа  
Направление подготовки 09.03.01 Информатика и вычислительная техника  
Направленность (профиль) «Технологии разработки программного обеспечения»  
форма обучения – очная

### **Курсовая работа**

по дисциплине «Пакеты прикладных программ для статистической обработки и  
анализа данных»

Оценка корреляции фактора рождаемости с помощью платформы анализа  
данных KNIME

Обучающейся 3 курса  
Нечаевой Натальи Андреевны

---

Руководитель:  
д.п.н, профессор  
Власова Е.З.

---

« \_\_\_\_\_ » \_\_\_\_\_ 2024 г.

Санкт-Петербург  
2024

## ОГЛАВЛЕНИЕ

<b>Введение .....</b>	<b>3</b>
<b>Основная часть .....</b>	<b>5</b>
<b>Теоретическая часть.....</b>	<b>5</b>
<b>Глава 1. Методологические основы и используемые термины .....</b>	<b>5</b>
<b>Глава 2. Реализация в Python .....</b>	<b>7</b>
<b>Практическая часть .....</b>	<b>11</b>
<b>Заключение.....</b>	<b>15</b>
<b>Источники.....</b>	<b>17</b>
<b>Приложение А .....</b>	<b>19</b>
<b>Приложение Б .....</b>	<b>20</b>
<b>Приложение В .....</b>	<b>22</b>

## Введение

В современном мире проблема низкого показателя рождаемости стала крайне актуальна. Снижение рождаемости может иметь серьезные последствия для демографической ситуации в стране, влияя на экономический рост, социальное обеспечение и устойчивость общества в целом. Важно выявить факторы, способствующие повышению рождаемости, а также оценить, как различные социальные и экономические условия могут влиять на этот показатель. В связи с этим, анализ взаимосвязей между рождаемостью, количеством браков и размером социальных выплат становится особенно важным для выработки эффективных мер государственной политики. [2][3]

Настоящая работа посвящена прогнозированию рождаемости в России. В рамках исследования будет проведен анализ данных о рождаемости, браках и социальных выплатах с 2000 по 2023 годы.

Предметом исследования курсовой работы является анализ корреляции фактора рождаемости от других факторов с использованием исторических данных средствами платформы для анализа данных KNIME.

Цель курсовой работы заключается в изучении возможностей платформы анализа данных KNIME при проведении корреляционного анализа и прогнозировании будущих показателей, а также в сравнении полученных данных с данными, полученными при работе с языком программирования Python.

Для достижения цели необходимо решить следующие задачи:

1. Изучить теоретические основы и алгоритмы реализации корреляционного анализа, провести анализ литературы и интернет-ресурсов по заданной теме.
2. Изучить возможности языка программирования Python для проведения корреляционного анализа и прогнозирования.

### 3. Изучить возможности платформы анализа данных KNIME для выполнения аналогичных задач.

Практическая значимость работы заключается в том, что полученные данные могут быть использованы для разработки государственных мер, направленных на повышение рождаемости.

Разработанный программный продукт может быть рекомендован для использования в аналитических центрах и государственных учреждениях, занимающихся вопросами демографии и социальной политики.

Структура курсовой работы включает в себя введение, основную часть, состоящую из теоретической части и практической части, заключение и список литературы.

Теоретическая часть работы посвящена основам корреляционного анализа, проведению исследований с помощью языка программирования Python.

Практическая часть работы посвящена реализации решения поставленной задачи в программе анализа данных KNIME.

Содержит 18 страниц, 12 рисунков, 3 таблицы, библиографический список, содержащий 12 источников, 3 приложения.

## **Основная часть**

### **Теоретическая часть**

#### **Глава 1. Методологические основы и используемые термины**

В данной работе используются данные о количестве браков и рождаемости на 1000 человек, а также размере единовременной социальной выплаты по рождению ребенка с 2000 по 2023 года. Таблица представлена в Приложении А.

Также нужно отметить, что размер социальной выплаты взят без региональных коэффициентов.

Данные были взяты из открытых источников. [10][11][12]

Корреляционный анализ – это метод обработки статистических данных, заключающийся в изучении коэффициентов корреляции между переменными. При этом сравниваются коэффициенты корреляции между одной парой или множеством пар признаков, для установления между ними статистических взаимосвязей.

Задача корреляционного анализа: определить тесноту связи между исследуемыми переменными. [6]

Теснота корреляционной связи определяется по коэффициенту корреляции ( $r$ ): сильная —  $r = \pm 0,7$  до  $\pm 1$ , средняя —  $r = \pm 0,3$  до  $\pm 0,699$ , слабая —  $r = 0$  до  $\pm 0,299$ . Данная классификация не является строгой.

Коэффициент корреляции рассчитывается по следующей формуле (1):

$$r_{xy} = \frac{\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}} \quad (1)$$

В данной работе исследуется корреляция между рождаемостью и двумя показателями: количеством браков и размером выплаты.

Для прогнозирования будет использоваться модель ARIMA (Autoregressive Integrated Moving Average) [1]. Она состоит из трех элементов:

1. AR (Autoregressive) – авторегрессионная часть модели, предполагает, что текущее значение временного ряда зависит от его предыдущих значений.
2. I (Integrated) – эта часть модели отвечает за стационарность временного ряда. Если временной ряд не является стационарным (например, имеет тренд или сезонность), то его необходимо дифференцировать.
3. MA (Moving Average) – часть модели, которая учитывает влияние случайных ошибок (шумов) на текущее значение временного ряда. Она предполагает, что текущее значение зависит от предыдущих ошибок предсказания.

Для проверки стационарности данных будет применяться тест Дики-Фуллера, который позволяет определить, являются ли временные ряды стационарными. Также будет использоваться критерий информационной оценки (AIC) для выбора наилучшей модели.

На основе имеющихся данных было выдвинуто две гипотезы:

1. Рождаемость зависит от количества браков.
2. Рождаемость зависит от размера социальной выплаты по рождению ребенка.

## Глава 2. Реализация в Python

Для построения модели прогнозирования ARIMA данные о рождаемости были необходимо было предварительно исследовать на статичность и выбрать наилучшие показатели. Для этого была реализована программа на языке программирования Python. Программа представлена в Приложении Б.

В программе используются наименования, представленные в Таблице 1:

Таблица 1. Наименования

Наименование	Значение
Births_per_1000	Название столбца, содержащего данные о рождаемости.
test_stationarity	Функция, проверяющая данные на стационарность. Использует метод Дика-Фуллера. Рассчитывает следующие параметры:  <ol style="list-style-type: none"><li>1. Статистика теста. Чем меньше это значение, тем более вероятно, что временной ряд стационарен.</li><li>2. р-значение. Если р-значение меньше 0.05, это указывает на то, что временной ряд стационарен.</li><li>3. Критические значения для различных уровней значимости (например, 1%, 5%, 10%).</li></ol>
timeseries	Проверяемые на стационарность данные
diff_count	Переменная, показывающая, сколько раз были совершены преобразования над исходными данными для достижения стационарности.
is_stationary	Переменная, которая хранит результат проверки стационарности временного ряда 'Births_per_1000'
best_aic	Переменная, которая хранит наименьшее значение AIC, найденное среди всех протестированных моделей.
Best_order	Переменная, которая хранит наилучшие параметры модели ARIMA (p, d, q), которые дают наименьшее значение AIC.

Разработанная программа позволяет выявить необходимые для прогнозирования параметры модели ARIMA. Результат работы программы представлен на Рисунке 1:

Лучшие параметры: (0, 2, 2) с AIC: 44.11579565360934

Рисунок 1. Лучшие параметры для модели ARIMA

Реализация поставленной цели с помощью языка программирования Python представлена в Приложении В.

Используемые в разработанной программе переменные указаны в Таблице 2:

Таблица 2. Наименования

Наименование	Значение
Year	Год, за который собраны данные
Marriages_per_1000	Количество браков на 1000 человек в каждом году.
Births_per_1000	Количество рождений на 1000 человек в каждом году.
Social_benefit	Социальная выплата по рождению ребенка в рублях.
correlation_marriages	Коэффициент корреляции между количеством браков и рождаемостью.
p_value_marriages	p-значение для теста корреляции между количеством браков и рождаемостью.
correlation_benefit	Коэффициент корреляции между социальными выплатами и рождаемостью.
p_value_benefit	p-значение для теста корреляции между социальными выплатами и рождаемостью.
alpha	Уровень значимости для проверки статистической значимости корреляции.

Результат работы программы представлен на Рисунках 2-6.



На Рисунке 2 представлены графики зависимости имеющихся данных о браках и рождаемости от года:

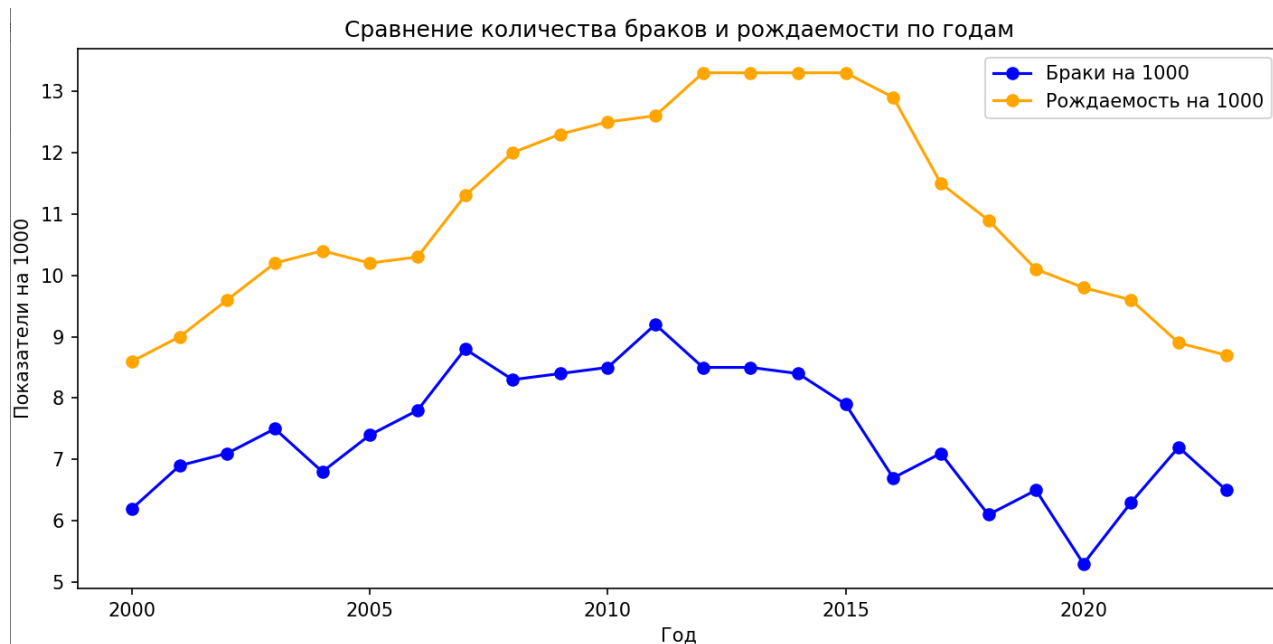


Рисунок 2. График браков и рождаемости

На Рисунке 3 представлено, как менялся размер социальной выплаты:



Рисунок 3. Изменение размера социальных выплат по годам

На Рисунке 4 представлены рассчитанные значения коэффициентов корреляции, а также р-значения:

```
Корреляция рождаемости от браков: 0.6857351148166607, р-значение: 0.00021678185569908904
Корреляция рождаемости от соц выплат: 0.12435100858104352, р-значение: 0.5626386652525772
Корреляция между рождаемостью и количеством браков статистически значима.
Корреляция между рождаемостью и социальными выплатами не значима статистически.
```

Рисунок 4. Результат вычислений

На Рисунке 5 и Рисунке 6 представлены результаты предсказаний рождаемости на 2024-2030 года. Для построения моделей прогноза ARIMA использовались значения, вычисленные ранее:

Прогноз рождаемости на 2024-2030 годы:		
	Year	Births
24	2024	8.417626
25	2025	8.053850
26	2026	7.690075
27	2027	7.326300
28	2028	6.962524
29	2029	6.598749
30	2030	6.234974

Рисунок 5. Предсказание рождаемости на 7 лет

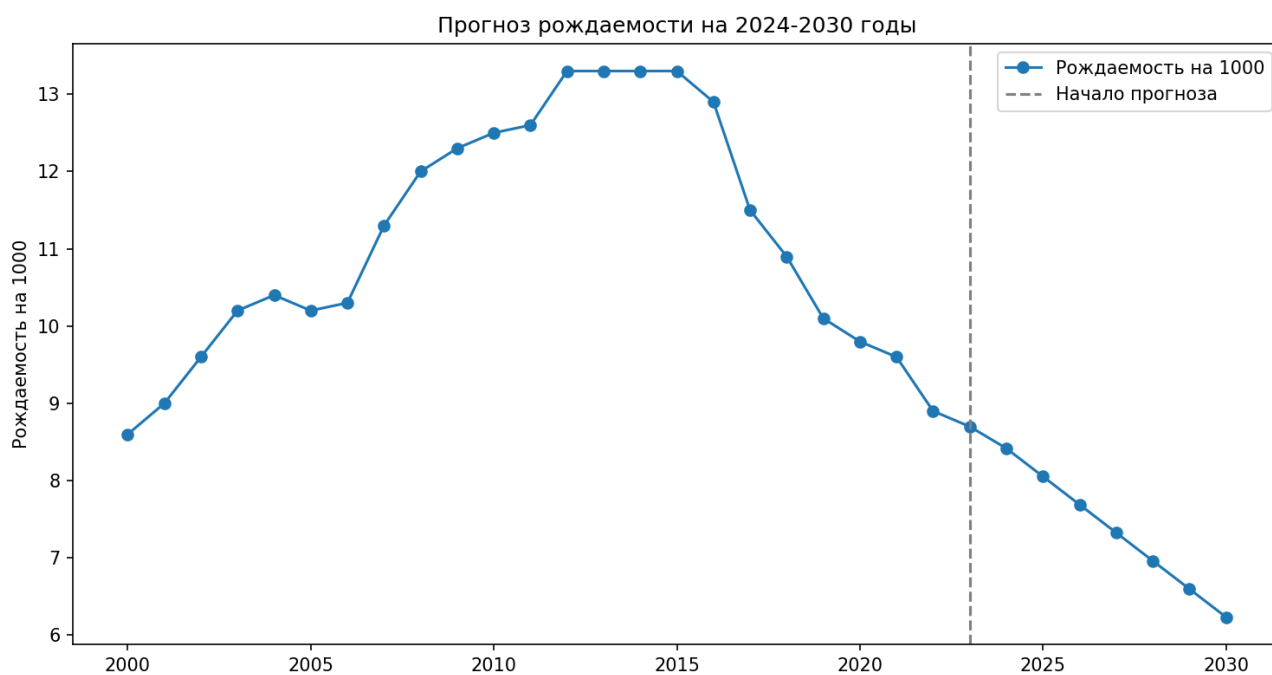


Рисунок 6. График рождаемости с предсказанными значениями

## Практическая часть

KNIME (Konstanz Information Miner) – это бесплатная платформа для анализа данных. []

В ней необходимо провести все те же вычисления, что и ранее в Python. Использоваться будут те же данные о браках, рождаемости и социальным выплатам по годам.

Для использования модели прогнозирования ARIMA, ее предварительно необходимо установить. Это можно сделать, проследовав по следующему маршруту в уже открытом workflow:

Menu → Install extensions → в открывшемся поисковом окне набрать “ARIMA” → выбрать в появившемся окне нужный пункт.

Разработанная структура представлена на Рисунке 7:

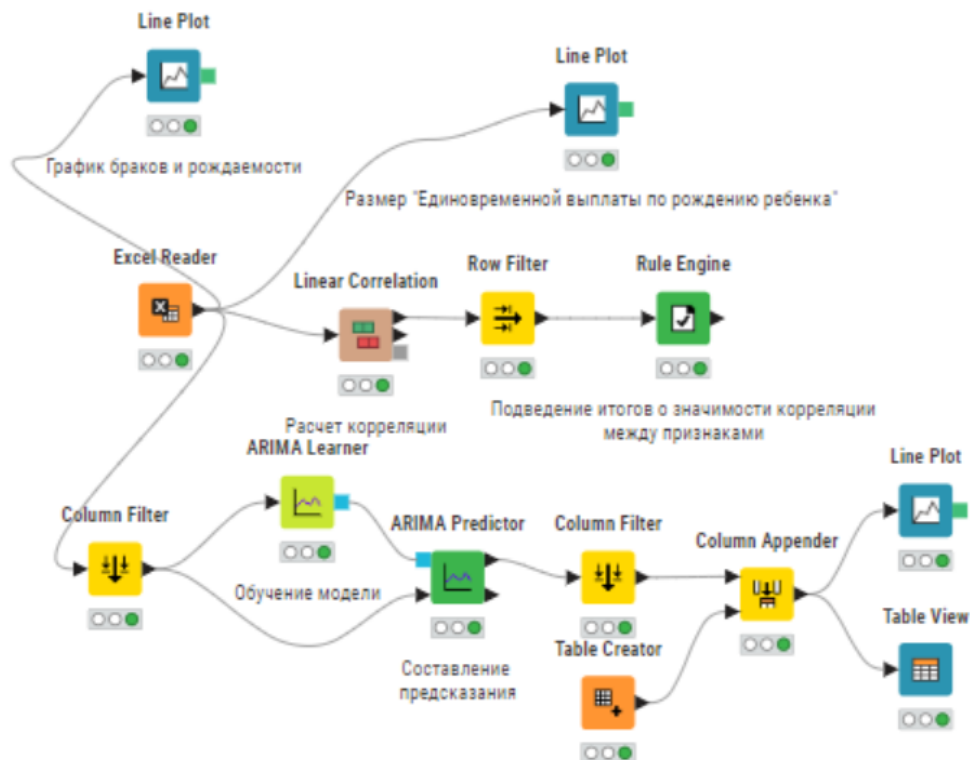


Рисунок 7. Анализ данных в KNIME

В работе были использованы следующие узлы:

1. Excel Reader – для извлечения данных из исходной таблицы.
2. Line Plot – для построения графиков.

3. Linear Correlation – для вычисления коэффициента корреляции и р-значения между факторами.
4. Row Filter – для фильтрации данных.
5. Rule Engine – для создания нового столбца таблицы, который задавался бы на основании введенных параметров. Для используемого в работе узла были заданы условия на Рисунке 8:

Expression		
\$	1	\$p value\$ <= 0.05 => "Significant"
\$	2	TRUE => "Not Significant"

Рисунок 8. Условия для узла Rule Engine

6. Column Filter – для фильтрации столбцов.
7. ARIMA Learner – для обучения модели ARIMA на основе имеющихся данных. Для нее были заданы условия (0, 2, 2) взятые в Главе 2. Реализация в Python.
8. ARIMA Predictor – для предсказания значений.
9. Table Creator – для создания таблицы, содержащей года с 2000 по 2030.
10. Column Appender – для объединения предсказанных значений и таблицы с годами.
11. Table View – для вывода таблицы с имеющимися данными и предсказанными по годам.

Графики по начальным данным представлены на Рисунках 9-10. Для красивого вывода данных в настройках были указаны названия легенд, шкал, а также заданы минимумы и максимумы по каждой шкале.

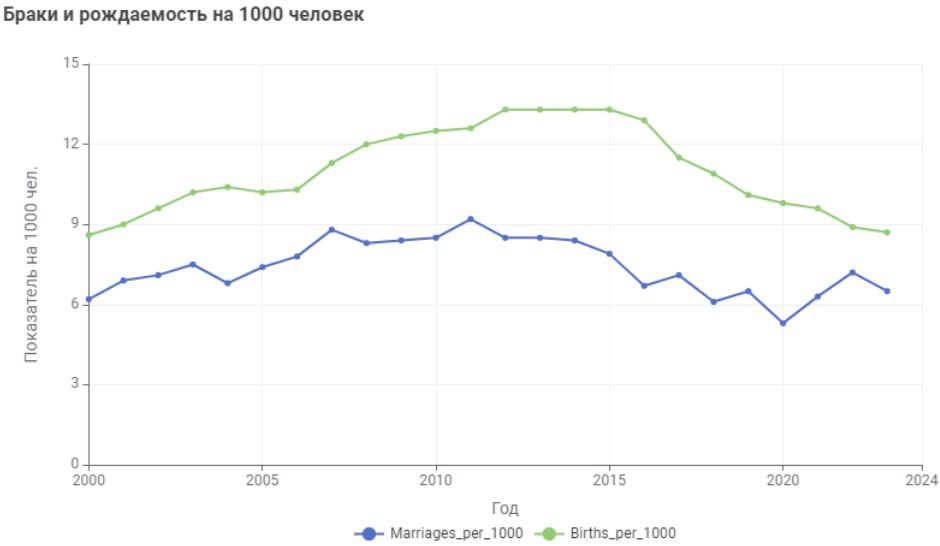


Рисунок 9. График браков и рождаемости с 2000 по 2023 года

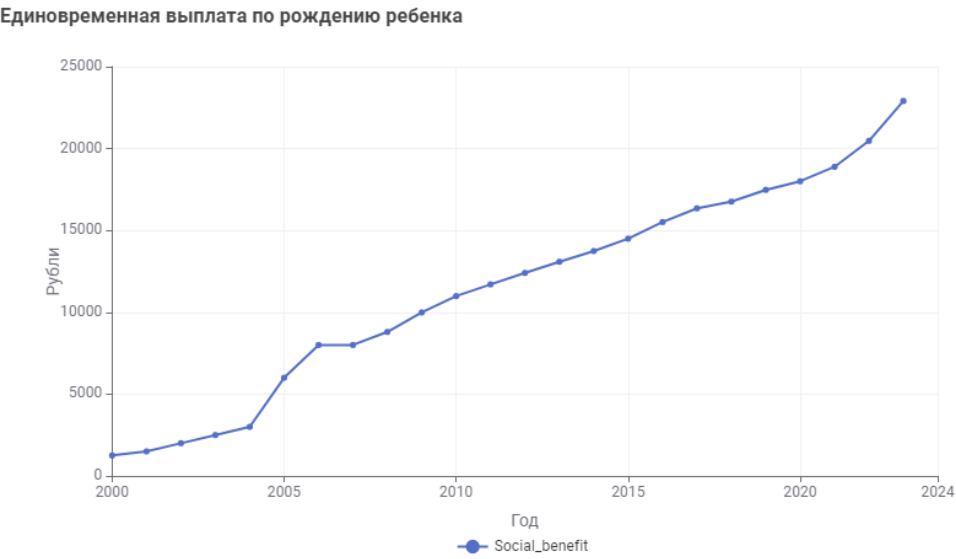


Рисунок 10. График социальных выплат с 2000 по 2023 года

Результат вычислений коэффициента корреляции и р-значения с кратким анализом полученных значений представлен на Рисунке 11:

Rows: 2 | Columns: 6

Table | Statistics

☐

#

RowID

First col...  
String

Second colu...  
String

Correlation ...  
Number (double)

p value  
Number (double)

Degrees of f...  
Number (integer)

prediction  
String

☐

1

Row0

Marriages\_per\_1...

Births\_per\_1000

0.686

0

22

Significant

☐

2

Row2

Births\_per\_1000

Social\_benefit

0.124

0.563

22

Not Significant

Рисунок 11. Результат вычислений

Результаты предсказаний, объединенные с уже имеющимися данными, представлены на Рисунке 12:

Рождаемость с 2000 по 2030 год (предсказание)

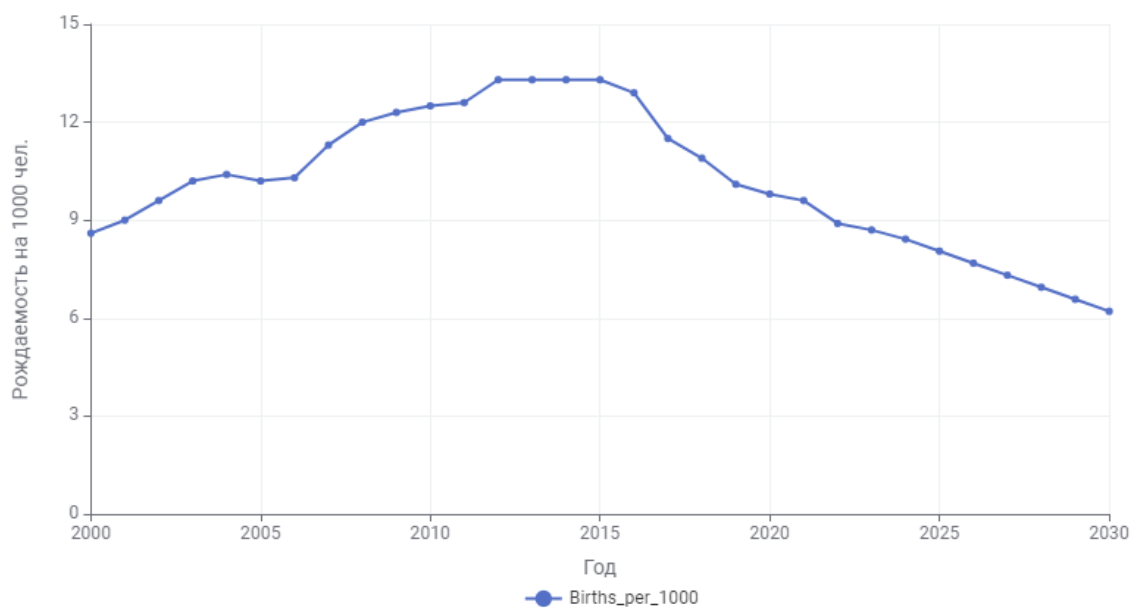


Рисунок 12. Рождаемость с 2000 по 2030 года  
Также можно эти данные увидеть в виде Таблицы 2:

Таблица 2. Предсказанные значения рождаемости на ближайшие годы

Предсказанное значение	Год
8.419242394996818	2024
8.050842804552238	2025
7.682443214107657	2026
7.314043623663077	2027
6.945644033218496	2028
6.577244442773916	2029
6.208844852329335	2030

## Заключение

В процессе выполнения курсовой работы была проведена аналитическая работа по исследованию зависимостей рождаемости от количества браков и размеров социальных выплат. На основе вычисленных корреляционных коэффициентов и р-значений, полученных как с использованием языка программирования Python, так и в среде KNIME, были сделаны следующие выводы:

1. Первая гипотеза, утверждающая о наличии статистически значимой связи между рождаемостью и количеством браков, была подтверждена. Это свидетельствует о том, что увеличение числа браков может положительно сказываться на уровне рождаемости, что подчеркивает важность укрепления института семьи в социальной политике.
2. Вторая гипотеза, предполагающая наличие статистически значимой связи между рождаемостью и размерами социальных выплат, была опровергнута. Это может говорить о том, что размер социальных выплат не является решающим фактором для повышения уровня рождаемости, и необходимо рассматривать более комплексные меры, направленные на поддержку семей.

Анализ построенных графиков показал, что существует тенденция к снижению как числа браков, так и уровня рождаемости, что является тревожным сигналом для демографической ситуации в стране.

Кроме того, на основе составленного прогноза о рождаемости в России на 2024-2030 годы можно сделать вывод, что если не предпринять эффективные меры по стимулированию рождаемости, то в будущем ситуация может ухудшиться. Это подчеркивает необходимость разработки и внедрения комплексных программ, направленных на поддержку семей, улучшение условий для воспитания детей и повышение уровня жизни.

Таким образом, результаты данного исследования подчеркивают важность анализа демографических тенденций и необходимость активных действий со стороны государства для улучшения ситуации в области рождаемости.



## Источники

[1] Бабёнышев, С. В., & Матеров, Е. Н. (2018). Математические методы и информационные технологии в научных исследованиях: учебное пособие. Железногорск: СПСА. 215 с. Текст: электронный. Лань: электронно-библиотечная система. URL: <https://e.lanbook.com/book/170699> (дата обращения: 10.10.2024). Режим доступа: для авториз. пользователей.

[2] Карманов, М. В., Егорова, Е. А., & Тузов, Д. А. (2009). Современные проблемы рождаемости населения в России. Статистика и экономика, (4). URL: <https://cyberleninka.ru/article/n/sovremennye-problemy-rozhdaemosti-naseleniya-v-rossii> (дата обращения: 11.10.2024).

[3] Макаренцева, А. О., Казенин, К. И., Середкина, Е. А., & Макаров, С. А. (2022). Рождаемость и поддержка рождаемости в России и мире в период пандемии коронавируса: доклад. Москва: Дело РАНХиГС. 90 с. ISBN 978-5-85006-395-5. Текст: электронный. Лань: электронно-библиотечная система. URL: <https://e.lanbook.com/book/293075> (дата обращения: 15.11.2024). Режим доступа: для авториз. пользователей.

[4] Население России 2019: двадцать седьмой ежегодный демографический доклад: доклад. (2022). Москва: Высшая школа экономики. 344 с. ISBN 978-5-7598-2438-1. Текст: электронный. Лань: электронно-библиотечная система. URL: <https://e.lanbook.com/book/259697> (дата обращения: 9.10.2024). Режим доступа: для авториз. пользователей.

[5] Пальмов, С. В. Основы сбора и обработки больших данных : учебное пособие / С. В. Пальмов. — Самара : ПГУТИ, 2023. — 285 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/411830> (дата обращения: 22.12.2024). — Режим доступа: для авториз. пользователей.

[6] Щурин, К. В. Надежность машин : учебное пособие / К. В. Щурин. — Санкт-Петербург : Лань, 2022. — 592 с. — ISBN 978-5-8114-3748-1. — Текст :

электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/206744> (дата обращения: 22.12.2024). — Режим доступа: для авториз. пользователей.

[7] Фахрисламова, Р. Т. (2014). Феномен женской бездетности в России: исторические и социальные предпосылки. Социологический журнал, (3). URL: <https://cyberleninka.ru/article/n/fenomen-zhenskoi-bezdetnosti-v-rossii-istoricheskie-i-sotsialnye-predposylki> (дата обращения: 09.12.2024).

[8] Чيو, К., & Фримэн, Д. (2020). Машинное обучение и безопасность: руководство (А. В. Снастин, Перев.). Москва: ДМК Пресс. 388 с. ISBN 978-5-97060-713-8. Текст: электронный. Лань: электронно-библиотечная система. URL: <https://e.lanbook.com/book/131707> (дата обращения: 09.12.2024). Режим доступа: для авториз. пользователей.

[9] Валерий Елизаров: Как менялись пособия на детей / Валерий Елизаров [Электронный ресурс] // Демоскоп : [сайт]. — URL: <https://www.demoscope.ru/weekly/2002/051/polit03.php> (дата обращения: 20.12.2024).

[10] Единовременное пособие при рождении ребенка / [Электронный ресурс] // КонсультантПлюс : [сайт]. — URL: [https://www.consultant.ru/document/cons\\_doc\\_LAW\\_83400/b5db1cba6790fba52a5de88b2ea70faaf25f7fbd/](https://www.consultant.ru/document/cons_doc_LAW_83400/b5db1cba6790fba52a5de88b2ea70faaf25f7fbd/) (дата обращения: 20.12.2024).

[11] Статистика браков и разводов в России по годам (Таблица) / [Электронный ресурс] // infotables : [сайт]. — URL: <https://infotables.ru/statistika/31-rossijskaya-federatsiya/787-statistika-brakov-i-razvodov> (дата обращения: 20.12.2024).

[12] Demographics of Russia / [Электронный ресурс] // Wikipedia : [сайт]. — URL: [https://en.wikipedia.org/wiki/Demographics\\_of\\_Russia](https://en.wikipedia.org/wiki/Demographics_of_Russia) (дата обращения: 10.10.2024).

## Приложение А

Год	Браки на 1000 человек	Рождаемость на 1000 человек	Социальные выплаты, рубли
2000	6,2	8,6	1252,0
2001	6,9	9,0	1500,0
2002	7,1	9,6	2000,0
2003	7,5	10,2	2500,0
2004	6,8	10,4	3000,0
2005	7,4	10,2	6000,0
2006	7,8	10,3	8000,0
2007	8,8	11,3	8000,0
2008	8,3	12,0	8800,0
2009	8,4	12,3	9989,0
2010	8,5	12,5	10988,0
2011	9,2	12,6	11703,0
2012	8,5	13,3	12405,0
2013	8,5	13,3	13087,0
2014	8,4	13,3	13741,0
2015	7,9	13,3	14497,0
2016	6,7	12,9	15512,0
2017	7,1	11,5	16350,0
2018	6,1	10,9	16759,0
2019	6,5	10,1	17479,0
2020	5,3	9,8	18004,0
2021	6,3	9,6	18886,0
2022	7,2	8,9	20472,0
2023	6,5	8,7	22909,0

## Приложение Б

```
1. import pandas as pd
2. import numpy as np
3. from statsmodels.tsa.stattools import adfuller
4. from statsmodels.tsa.arima.model import ARIMA
5.
6. # Функция для проверки стационарности
7. def test_stationarity(timeseries):
8.     result = adfuller(timeseries)
9.     print('Статистика теста:', result[0])
10.    print('p-значение:', result[1])
11.    print('Критические значения:')
12.    for key, value in result[4].items():
13.        print(f' {key}: {value}')
14.
15.    return result[1] < 0.05 # Возвращаем True, если ряд стационарен
16.
17. # Загрузка данных
18. file_path = 'path' # Путь к файлу с данными
19. df = pd.read_excel(file_path)
20.
21. # Проверка стационарности временного ряда 'Births_per_1000'
22. print("Проверка стационарности временного ряда 'Births_per_1000':")
23. is_stationary = test_stationarity(df['Births_per_1000'])
24.
25. diff_count = 0 # Счетчик разностей
26. print('Ищем количество разностей:')
27. while not is_stationary:
28.     # Применяем разности
29.     df['Births_per_1000'] = df['Births_per_1000'].diff().dropna()
30.     df = df.dropna() # Удаляем строки с NaN
31.     diff_count += 1 # Увеличиваем счетчик разностей
32.
33.     # Проверяем стационарность
34.     print(f'Применены {diff_count} разности к временному ряду 'Births_per_1000'.")
35.     print("Проверка стационарности разностей 'Births_per_1000':")
36.     is_stationary = test_stationarity(df['Births_per_1000'])
37.
38. # После завершения цикла
39. print(f'Общее количество примененных разностей: {diff_count}')
40.
41. # Определяем количество наблюдений
42. n = len(df['Births_per_1000'])
43.
44. # Пример оценки модели ARIMA с различными p и q
45. best_aic = np.inf
46. best_order = None
47. best_model = None
48.
49. # Пробуем разные значения p и q
50. for p in range(5): # Измените диапазон по необходимости
51.     for q in range(5): # Измените диапазон по необходимости
```

```

52.     try:
53.         model = ARIMA(df['Births_per_1000'], order=(p, diff_count, q))
54.         model_fit = model.fit()
55.         print(f'ARIMA({p}, {diff_count}, {q}) - AIC: {model_fit.aic}')
56.         if model_fit.aic < best_aic:
57.             best_aic = model_fit.aic
58.             best_order = (p, diff_count, q)
59.             best_model = model_fit
60.     except Exception as e:
61.         print(f'Ошибка при оценке модели ARIMA({p}, {diff_count}, {q}): {e}')
62.         continue
63.
64. print(f'Лучшие параметры: {best_order} с AIC: {best_aic}')

```

## Приложение В

```
1. import pandas as pd
2. import matplotlib.pyplot as plt
3. from scipy.stats import pearsonr
4. from statsmodels.tsa.arima.model import ARIMA
5.
6. # Шаг 1: Чтение данных из файла Excel
7. file_path = ('path') # Путь к файлу с данными
8. df = pd.read_excel(file_path)
9.
10. # 2. Построение графиков
11. plt.figure(figsize=(12, 6))
12.
13. # График количества браков
14. plt.plot(df['Year'], df['Marriages_per_1000'], marker='o', label='Браки на 1000',
           color='blue')
15.
16. # График рождаемости
17. plt.plot(df['Year'], df['Births_per_1000'], marker='o', label='Рождаемость на 1000',
           color='orange')
18.
19. # Настройка меток и заголовка
20. plt.title('Сравнение количества браков и рождаемости по годам')
21. plt.xlabel('Год')
22. plt.ylabel('Показатели на 1000')
23. plt.legend() # Добавление легенды
24. plt.tight_layout()
25. # Отображение графика
26. plt.show()
27.
28. plt.subplot(1,1,1)
29. plt.plot(df['Year'], df['Social_benefit'], marker='o', color='green')
30. plt.title('Социальная выплата по рождению ребенка по годам')
31. plt.xlabel('Год')
32. plt.ylabel('Выплата (рубли)')
```

```

33. plt.tight_layout()
34. plt.show()
35.
36. # 3. Нахождение корреляции
37. correlation_marriages, p_value_marriages = pearsonr(df['Marriages_per_1000'],
    df['Births_per_1000'])
38. correlation_benefit, p_value_benefit = pearsonr(df['Social_benefit'],
    df['Births_per_1000'])
39.
40. print(f'Корреляция рождаемости от браков: {correlation_marriages}, p-значение:
    {p_value_marriages}')
41. print(f'Корреляция рождаемости от соц выплат: {correlation_benefit}, p-значение:
    {p_value_benefit}')
42. # Интерпретация p-значений
43. alpha = 0.05 # уровень значимости
44. if p_value_marriages < alpha:
45.     print("Корреляция между рождаемостью и количеством браков статистически
        значима.")
46. else:
47.     print("Корреляция между рождаемостью и количеством браков не значима
        статистически.")
48.
49. if p_value_benefit < alpha:
50.     print("Корреляция между рождаемостью и социальными выплатами
        статистически значима.")
51. else:
52.     print("Корреляция между рождаемостью и социальными выплатами не значима
        статистически.")
53.
54. # 4. Прогнозирование
55. births_model = ARIMA(df['Births_per_1000'], order=(0, 2, 2)).fit()
56. # Прогнозируем на 2024-2030 годы (7 лет)
57. future_births = births_model.forecast(steps=7)
58.
59. # Вывод результатов

```

```

60. years_forecast = list(range(2024, 2031)) # Годы с 2024 по 2030
61. forecast_results = pd.DataFrame({'Year': years_forecast, 'Births': future_births})
62.
63. print("Прогноз рождаемости на 2024-2030 годы:")
64. print(forecast_results)
65.
66. # Объединяем исторические данные и прогнозы для графика
67. new_df = pd.DataFrame({'Year': df['Year'], 'Births': df['Births_per_1000']})
68. full_df = pd.concat([new_df, forecast_results])
69.
70. # Дополнение графика
71. plt.figure(figsize=(12, 6))
72. plt.plot(full_df['Year'], full_df['Births'], marker='o', label='Рождаемость на 1000')
73. plt.title('Прогноз рождаемости на 2024-2030 годы')
74. plt.xlabel('Год')
75. plt.ylabel('Рождаемость на 1000')
76. plt.axvline(x=2023, color='grey', linestyle='--', label='Начало прогноза')
77. plt.legend()
78. plt.show()

```