**Домашнее творческое задание по теме: «Исследование и прогнозирование показателей, характеризующих демографическое положение в Российской Федерации.»**

**Цель работы:** Исследование демографических показателей Российской Федерации и их прогнозирование.

**Актуальность:**

Исследование демографических показателей и прогнозирование демографической ситуации – одна из важнейших задач для Российской Федерации. В последние десятилетия Россия сталкивается с рядом серьезных демографических вызовов, среди которых особенно остро стоят низкая рождаемость, высокая смертность (прежде всего в трудоспособном возрасте), старение населения и отток населения из регионов. Эти негативные тенденции могут привести к сокращению численности населения, дефициту трудовых ресурсов, увеличению нагрузки на пенсионную систему и другим социально-экономическим проблемам.

Осознавая серьезность сложившейся ситуации, государство уделяет пристальное внимание решению демографических проблем. В 2019 году был запущен национальный проект "Демография", направленный на повышение рождаемости, снижение смертности, увеличение продолжительности жизни и улучшение миграционной ситуации.

**Введение:**

В рамках данной работы проведено исследование ключевых демографических показателей Российской Федерации за период с 1991 по 2022 года. Для анализа были использованы данные с сайта rosstat.gov.ru, включающие следующие показатели: Коэффициент браков на 1000 чел, Коэффициент рождаемости, Коэффициент смертности, Коэффициент миграции, Заболеваемость, Удельный вес населения старше трудоспособного возраста, Население в трудоспособном возрасте, Численность экономически активного населения, Безработица, Средняя номинальная зарплата, Инвестиции в жилое строительство, Квадратные метры жилья на 1 человека, Охват дошкольным образованием, Загрязнение окружающей среды.



Работа состоит из следующих частей: обзор данных, определение зависимых и независимых переменных, построение и анализ моделей линейной регрессии, спецификация и проверка идентифицируемости модели, оценка модели методом трехшаговых наименьших квадратов (3МНК), прогнозирование, заключение.

**Определение зависимых и независимых переменных для построения моделей:**

Для определения взаимосвязей между показателями была построена матрица корреляции, на основе которой были выделены зависимые (эндогенные) и независимые (экзогенные) переменные.



Сильные положительные корреляции:

* + "населения в трудоспособном возрасте" и "численность экономически активного населения" (0.95) - логично, чем больше людей в трудоспособном возрасте, тем больше экономически активных.
  + "Коэф. Рожд" и "Коэф. Миграц" (0.8) - возможно, рост рождаемости делает страну более привлекательной для мигрантов.

Сильные отрицательные корреляции:

* + "удельный вес населения старше трудоспособного возраста" и "населения в трудоспособном возрасте" (-0.94) - естественная зависимость, чем больше пожилых, тем меньше людей в трудоспособном возрасте.

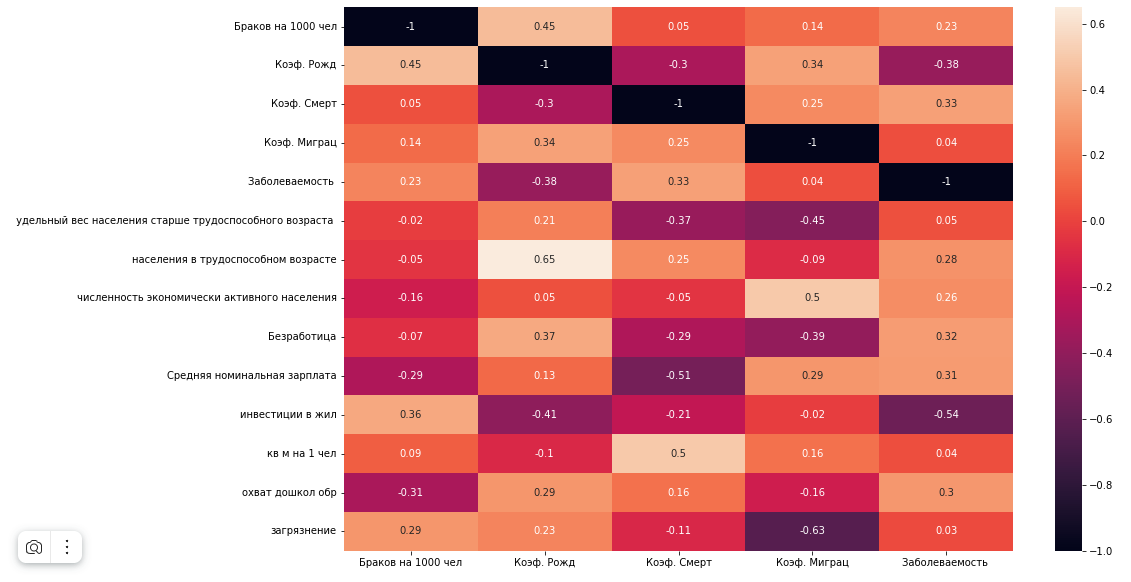
Умеренные корреляции:

* + "Браков на 1000 чел." и "Коэф. Рожд" (0.72) - ожидаемая связь, больше браков, потенциально больше рождений.
  + "Заболеваемость" и "удельный вес населения старше трудоспособного возраста" (0.76) - заболеваемость часто выше среди пожилого населения.

Слабые корреляции:

* + Много переменных имеют слабую корреляцию, что говорит об отсутствии прямой линейной зависимости между ними

Для более точного определения взаимосвязей была построена матрица частных коэффициентов корреляции.



На основе анализа корреляционных матриц были выбраны следующие зависимые и независимые переменные:

Зависимые переменные:

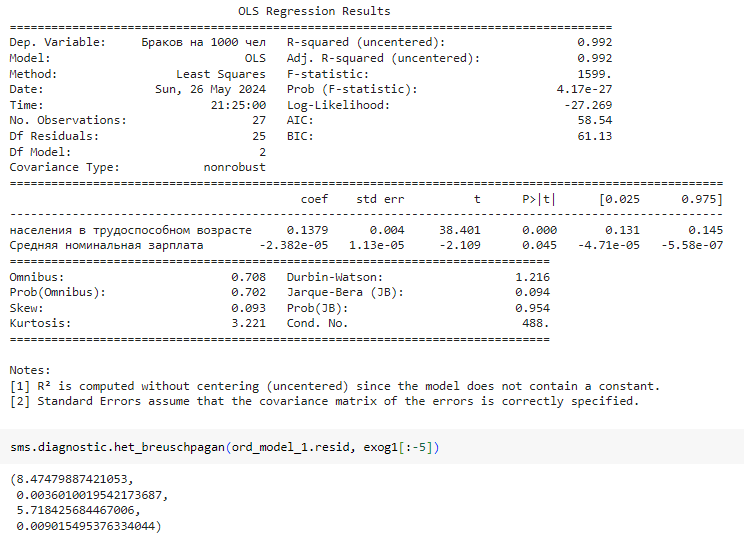
* Y1: Браков на 1000 чел
* Y2: Коэффициент рождаемости
* Y3: Коэффициент смертности
* Y4: Коэффициент миграции
* Y5: Заболеваемость

Независимые переменные:

* X1: Удельный вес населения старше трудоспособного возраста
* X2: Население в трудоспособном возрасте
* X3: Численность экономически активного населения
* X4: Безработица
* X5: Средняя номинальная зарплата
* X6: Инвестиции в жилое строительство
* X7: Квадратные метры жилья на 1 человека
* X8: Охват дошкольным образованием
* X9: Загрязнение окружающей среды

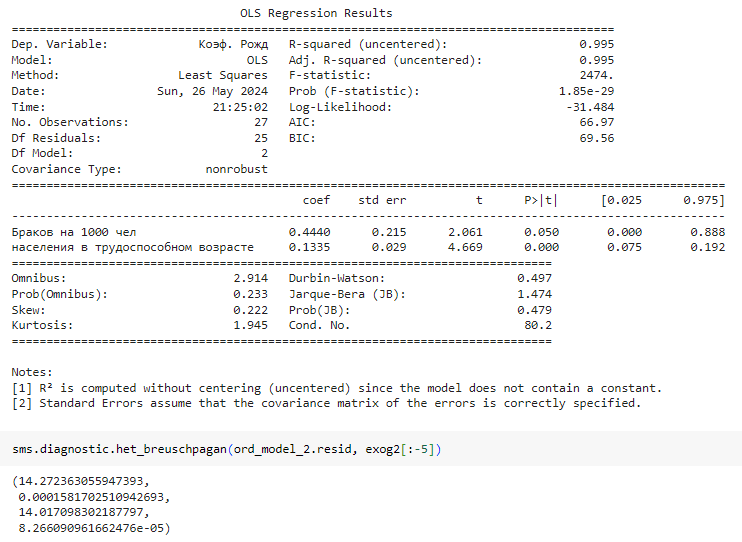
**На основе данных о корреляции были построены модели линейной регрессии**

Для первой модели были взяты переменные: Y1: Браков на 1000 чел., X2: Население в трудоспособном возрасте и X5: Средняя номинальная зарплата. Ожидается, что на количество браков будут влиять такие факторы, как численность населения в трудоспособном возрасте и уровень доходов, отражающий экономическую ситуацию в стране.



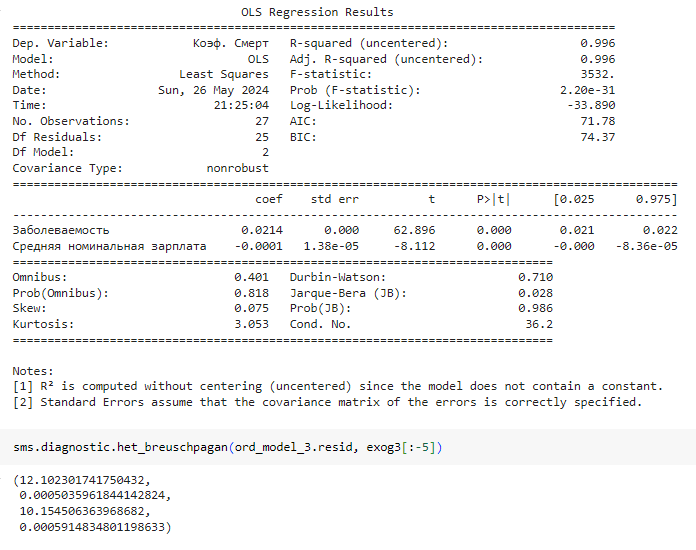
Эта модель показала, что параметры и уравнение значимы, остатки распределены нормально и наблюдается положительная автокорреляция. Также демонстрирует статистическую значимость, что подтверждается высокими значениями коэффициента детерминации (R-квадрат

Для второй модели были взяты переменные: Y2: Коэффициент рождаемости, • X2: Население в трудоспособном возрасте и Y1: Браков на 1000 чел. Предполагается, что рождаемость связана с количеством браков, а также с численностью населения в трудоспособном возрасте.



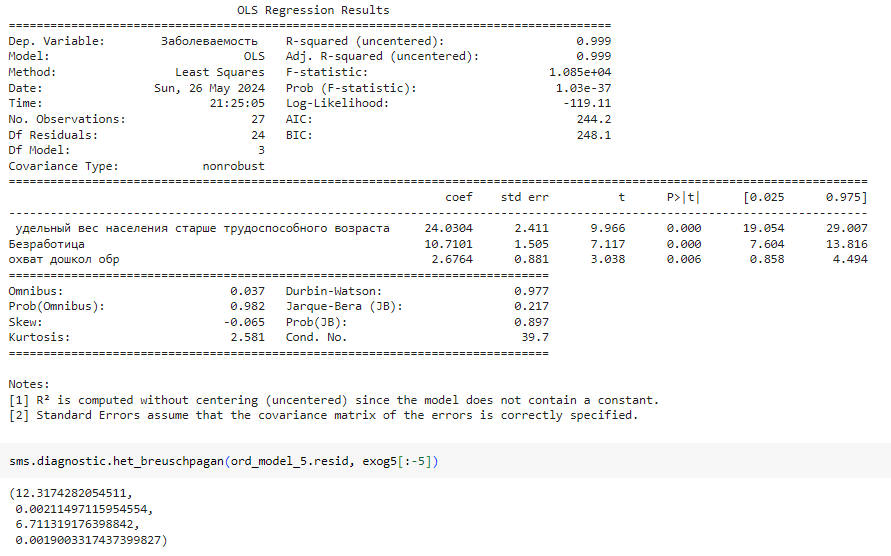
Вторая модель показала, что параметры и уравнение значимы, остатки распределены нормально и наблюдается положительная автокорреляция. Результаты нашего анализа показывают, что построенная модель обладает статистической значимостью, что отражается в высоких значениях R-квадрат.

Для третей модели были взяты переменные: Y3: Коэффициент смертности, X5: Средняя номинальная зарплата и Y5: Заболеваемость. Предполагается, что смертность связана с уровнем доходов, который может влиять на качество здравоохранения, и с уровнем заболеваемости.



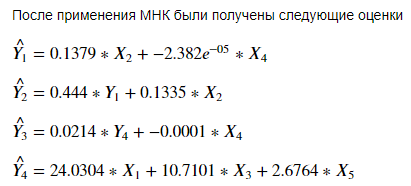
Третья модель показала, что параметры и уравнение значимы, остатки распределены нормально и наблюдается положительная автокорреляция. Высокие значения R-квадрат, полученные в результате построения модели, указывают на ее статистическую значимость.

Для пятой модели были взяты переменные: Y5: Заболеваемость, X1: Удельный вес населения старше трудоспособного возраста, X4: Безработица, X8: Охват дошкольным образованием. Ожидается, что на уровень заболеваемости влияют такие факторы, как удельный вес населения старше трудоспособного возраста, уровень безработицы и охват дошкольным образованием, который может косвенно отражать уровень санитарной культуры.

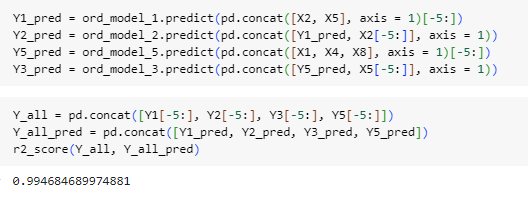


Пятая модель показала, что параметры и уравнение значимы, остатки распределены нормально и наблюдается положительная автокорреляция. На основе высоких значений R-квадрат можно сделать вывод о статистической значимости разработанной модели.

**Вывод по моделям:**



Все модели получились значимыми, кроме моделей для коэффициента миграции, поэтому было решено его не учитывать.



R2 = 0.9946 на тестовой выборке, что говорит о хорошей предсказательной способности модели.

В ходе работы были получены следующие переменные:

Зависимые (целевые) переменные

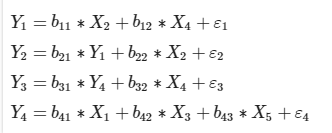
* Y1: Браков на 1000 чел
* Y2: Коэффициент рождаемости
* Y3: Коэффициент смертности
* Y4: Заболеваемость

Независимые переменные

* X1: Удельный вес населения старше трудоспособного возраста
* X2: Население в трудоспособном возрасте
* X3: Безработица
* X4: Средняя номинальная зарплата
* X5: Охват дошкольным образованием

**Спецификация модели и проверка идентифицируемости модели.**

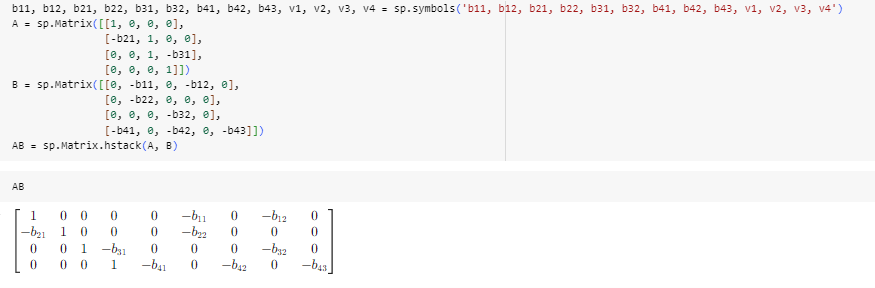
На основе анализа корреляции и линейных моделей была предложена следующая спецификация модели системы одновременных уравнений:

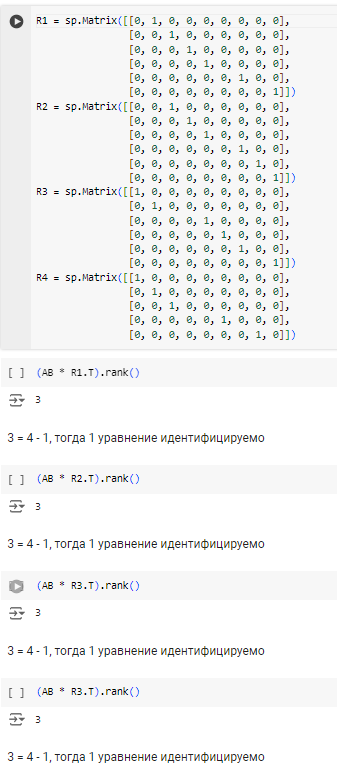


**Проверка идентифицируемости:**

Для проверки идентифицируемости модели были использованы метод порядка и матричный метод (ранг).

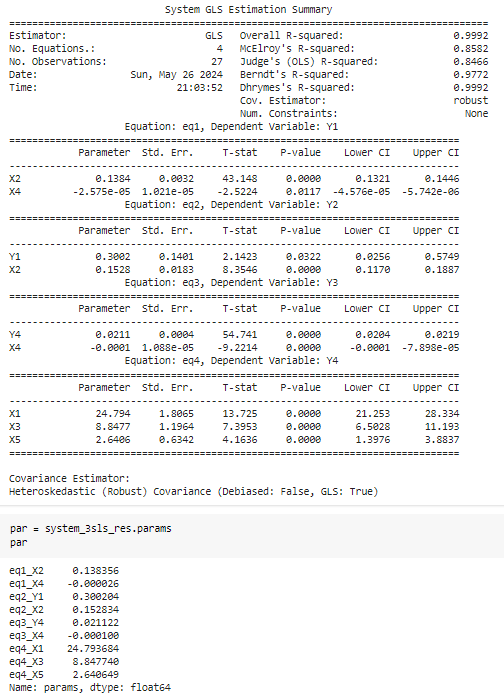
* Метод порядка:
* Матричный метод (ранг):



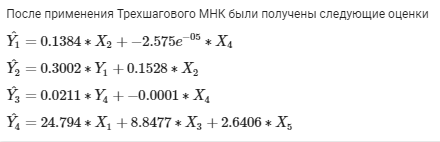


**Трехшаговый метод наименьших квадратов:**

Для оценки параметров модели системы одновременных уравнений был выбран трехшаговый метод наименьших квадратов (3МНК). Этот метод позволяет получить состоятельные и эффективные оценки параметров в случае системы одновременных уравнений, где присутствуют эндогенные объясняющие переменные.



Для всех параметров p-value t статистик > 0.05, значит H0 отвергается, следовательно параметры значимы. R2 = 0.9992, что говорит об хорошем качестве модели.



Предсказания модели 3МНК были оценены и был получен R2 = 0.9949 на тестовой выборке, что говорит о хорошей предсказательной способности модели.

**Вывод**

В ходе исследования было выявлено, что на коэффициент брачности сильнее всего влияет доля населения в трудоспособном возрасте и средняя номинальная зарплата, это связано с тем, что при более высоких зарплатах люди более уверены в завтрашнем дне и готовы формировать семьи и планировать семью, а население трудоспособного возраста более склонно вступать в брак. Для коэффициента рождаемости наиболее значимыми факторами оказались коэффициент брачности и доля трудоспособного населения, потому что именно люди это возраста чаще всего рожают детей и находясь в браке люди более склоны к тому, чтобы заводить детей. На коэффициент смертности большее влияние оказывают коэффициент заболеваемости и размер средней номинальной зарплаты, это следует из того, что многие болезни в ходе своего развития могут привести к осложнениям, ведущим к смерти, и люди с большим достатком имеют больше возможностей для ведения здорового образа жизни, раннего диагностирования и более продвинутого лечения болезней, что снижает смертность. Коэффициент заболеваемости населения сильнее всего зависит от доли населения старше трудоспособного возраста, уровня безработицы и охвата дошкольным образованием, потому что возрастные люди и дети наиболее уязвимые категории населения для болезней, а активное взаимодействие детей в организациях дошкольного обучения может являться фактором распространения некоторых заболеваний. Уровень безработицы может влиять на заболеваемость тем, что в кризисные моменты, когда безработица растёт, у людей становится меньше возможностей для того, чтобы следить за своим здоровьем, поэтому число заболевших растёт.

Основываясь на полученных результатах, можно сделать вывод о том, что правительству Российской Федерации стоит сосредоточиться на улучшении материального благосостояния населения, увеличении доли трудоспособного населения и формировании устойчивой экономической ситуации в стране. Так же государству следует обратить внимание на защиту пожилых людей и детей от воздействий заболеваний, проводя регулярную вакцинацию и напоминаний и методах индивидуальной защиты.