Федеральное государственное образовательное бюджетное учреждение высшего образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**(Финансовый университет)**

Департамент анализа данных и машинного обучения

ИТОГОВАЯ ДОМАШНЯЯ РАБОТА

по дисциплине «Эконометрика»

на тему:

**«Панельные данные»**

**Выполнил:**

Джаводов С. Д.

группа ПМ20-1

**Москва, 2023**

Содержание

**Введение………………………………………………...3**

**Задача 10………………………………………………...4**

**Задача 11-15………………………………………………...5-15**

**ВВЕДЕНИЕ**

С 2016 года в России начался второй этап депопуляции: уровень смертности вновь превысил уровень рождаемости, естественный прирост сменился убылью населения. Прогнозные оценки свидетельствуют о нарастании данного негативного тренда в ближайшие десятилетия. Отчасти потери населения вследствие естественной убыли компенсируются миграционным приростом, однако иммиграционный ресурс постепенно иссякает, что актуализирует вопрос изучения факторов, обусловливающих современные демографические тенденции, для понимания причин и определения возможностей управления ими. Целью данной работы стал анализ комплекса факторов разной природы и направленности их влияния на демографическое развитие России и ее регионов. Для реализации цели исследования использован метод регрессионного анализа панельных данных. Информационной базой послужили данные Федеральной службы государственной статистики за период с 2010 по 2017 год по 80 субъектам РФ. Зависимыми переменными выступили суммарный коэффициент рождаемости, ожидаемая продолжительность жизни при рождении, коэффициенты интенсивности по прибытию и выбытию.

## **ЗАДАЧА 10**

Информационной базой нашего исследования послужили данные Федеральной службы государственной статистики за период с 2010 по 2017 год по 80 субъектам РФ. Из анализа были исключены Республика Крым, г. Севастополь, Ямало-Ненецкий, Ханты-Мансийский и Ненецкий автономные округа из-за отсутствия или неполноты статистической информации по ряду показателей. На основе исследований Синяковской и Тындика, следует включить возраст женщины и брачный статус.[[1]](#footnote-1) На основе исследований Архангельского, Шабунова и Калачикова, следует включить занятость (будем рассматривать численность безработных в субъекте, как в статье[[2]](#footnote-2)), уровень дохода (номинальная и реальная ЗП, прожиточный минимум на детей, как в статье2). На основе статьи2 стоит включить обеспеченность амбулаторно-поликлиническими учреждениями и число прерывания беременности. Кроме того, на мой взгляд, стоит рассмотреть уровень оказания медицинских услуг, который можно оценить на основе материнской и младенческой смертности.

Тогда спецификация примет вид:

* x1 – средний возраст матери при рождении ребёнка
* x2 – общий коэффициент брачности
* x3 – численность безработных в возрасте 15-72 лет
* x4 – среднемесячная номинальная начисленная заработная плата работников
* x5 – реальная среднемесячная начисленная заработная плата работников
* x6 – величина прожиточного минимума на ребенка в возрасте до 16 лет
* x7 – число фельдшерско-акушерских пунктов
* x8 – число прерываний беременности
* x9 – материнская смертность
* x10 – коэффициент младенческой смертности
* y – общий коэффициент рождаемости

Визуализируем корреляционную матрицу для оценки тесноты связи зависимой переменной и регрессоров:

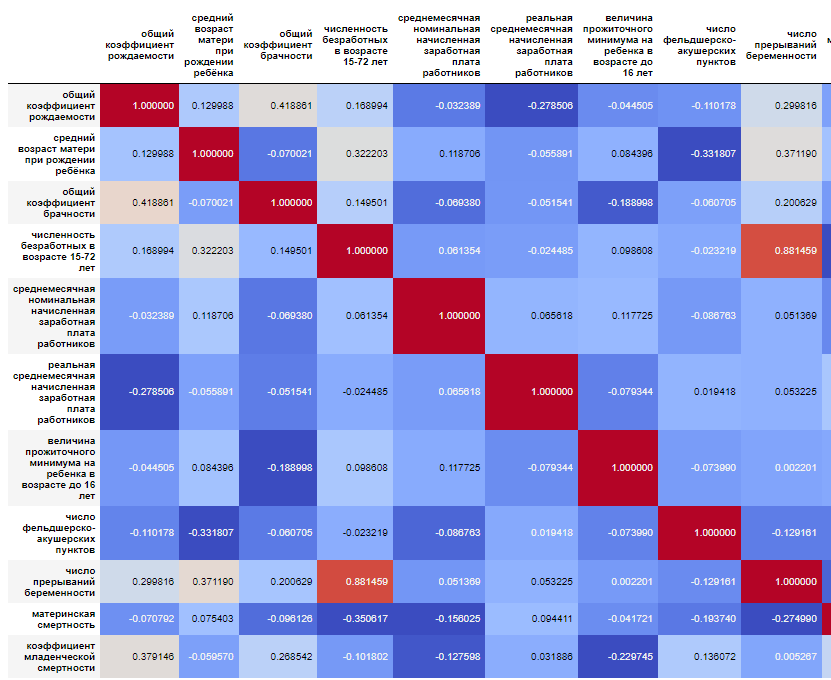


Рисунок 1. Корреляционная матрица

Также для каждого субъекта по отдельности строится корреляционная матрица. Ниже приведены корреляционные матрицы по некоторым субъектам (для полной информации обратитесь в pdf файл с расчётами):

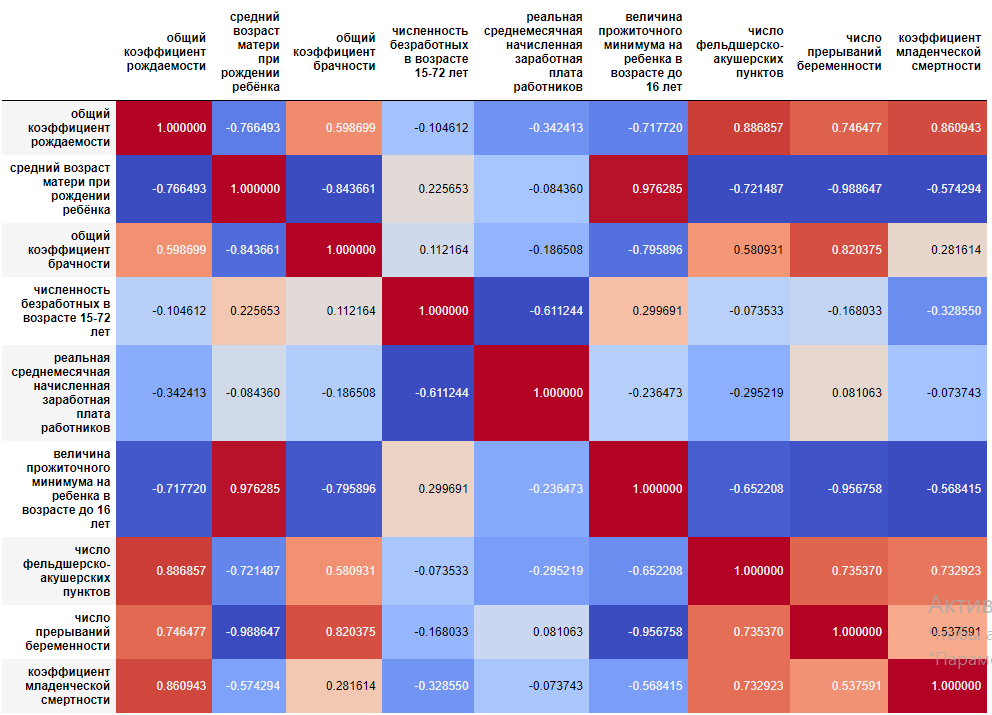


Рисунок 2. Корреляционная матрица (Белгородская область)

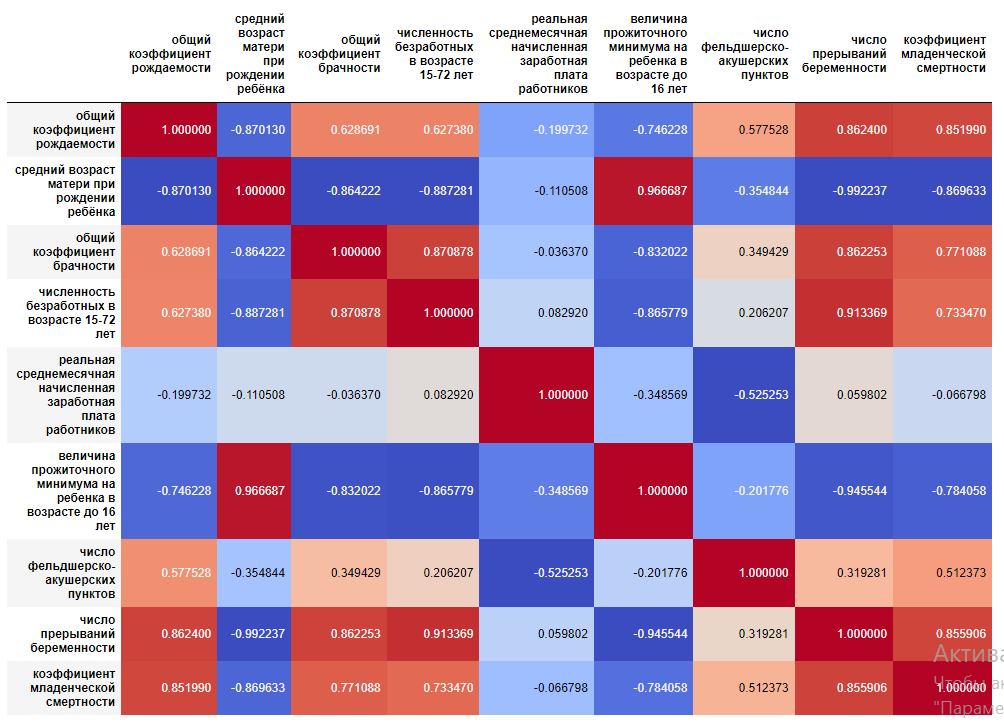


Рисунок 3. Корреляционная матрица (Брянская область)

Как видно из анализа, для различных субъектов разные факторы являются наиболее коррелированными с рождаемостью. Но среднемесячная номинальная начисленная заработная плата работников и материнская смертность дает небольшой вклад, к тому же, в столбце материнская смертность очень много пропусков по данным Росстата, которые нельзя корректно воспроизвести. Поэтому перечисленные выше факторы стоит исключить из дальнейшего рассмотрения.

# **ЗАДАЧА 11-15**

Модель сквозной (объединенной) регрессии (pooled model) предписывает одинаковое поведение всем объектам выборки во все моменты времени, она не учитывает индивидуальных различий, из-за чего является самой ограничительной из всех видов. Объединенная модель описывается уравнением:

Построив объеденную модель (на Python) на данных используемых в работе получили следующий результат:

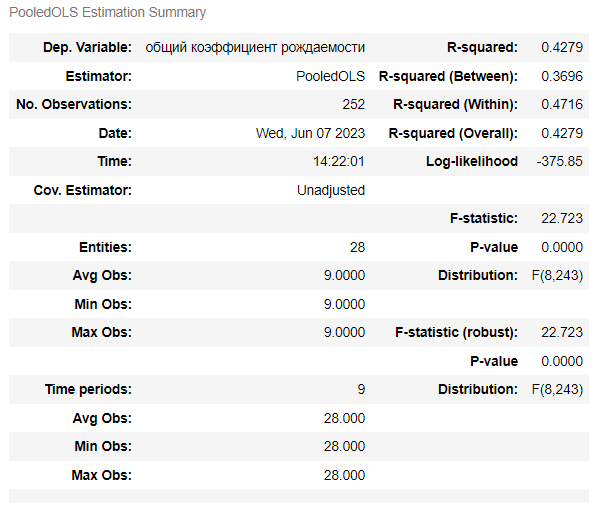




Рисунок 4. Объеденная модель

Коэффициенты при факторах 'средний возраст матери при рождении ребёнка’, ‘число фельдшерско-акушерских пунктов’, ‘величина прожиточного минимума на ребенка в возрасте до 16 лет' являются статистически незначимыми при значении p-критерия <0,01. Уберём данные факторы и повторно построим модель:





Рисунок 5. Объединённая модель (после удаления незначимых факторов)

Все коэффициенты статистически значимы при p <0,01. Запишем уравнение модели подставив коэффициенты:

Yit = 15.243 + 0.3056x1it -0.0219x2it -0.0811x3it + 0.0001x4it + 0.2406x5it + eit

Таблица 1. Меры качества подгонки

|  |  |
| --- | --- |
| R-squared(коэффициент детерминации) | 0,4169 |
| Межгрупповой R-squared(R-squared(between)) | 0,3208 |
| Внутригрупповой R-squared(R-squared(within)) | 0,4888 |

**Модель с фиксированными эффектами:**

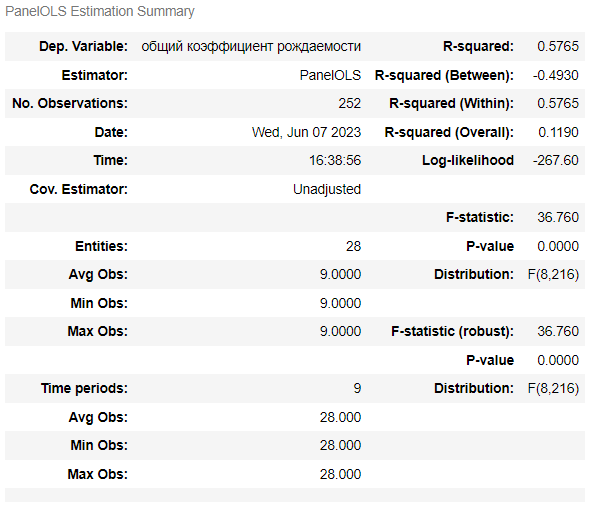




Рисунок 6. Модель с фиксированными эффектами

Коэффициенты при факторах 'число фельдшерско-акушерских пунктов', 'число прерываний беременности', 'общий коэффициент брачности', 'средний возраст матери при рождении ребёнка' являются статистически незначимыми при значении p-критерия <0,01. Уберём данные факторы и повторно построим модель:





Рисунок 7. Модель с фиксированными эффектами

x1 – численность безработных в возрасте 15-72 лет

x2 – реальная среднемесячная начисленная заработная плата работников

x3 – величина прожиточного минимума на ребенка в возрасте до 16 лет

x4–коэффициент младенческой смертности 𝑦𝑖𝑡=𝑎𝑖+0.0154𝑥1𝑖𝑡−0.0841𝑥2𝑖𝑡−0.0002𝑥3𝑖𝑡+0.2888𝑥4𝑖𝑡+𝑒𝑖𝑡, где 𝑎𝑖

- индивидуальный для i-й панель параметр (ниже приведено несколько значений)

Таблица 2. Меры качества подгонки

|  |  |
| --- | --- |
| R-squared overall | -0,0186 |
| Межгрупповой R-squared(R-squared(between)) | -0,7987 |
| Внутригрупповой R-squared(R-squared(within)) | 0,5446 |
| R-squared | 0,5446 |

**Модель со случайными эффектами:**

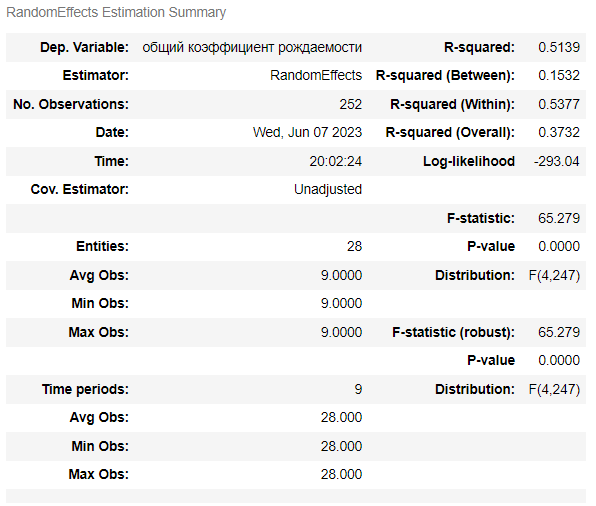




Рисунок8. Модель со случайными эффектами

x1 – общий коэффициент брачности

x2 – реальная среднемесячная начисленная заработная плата работников

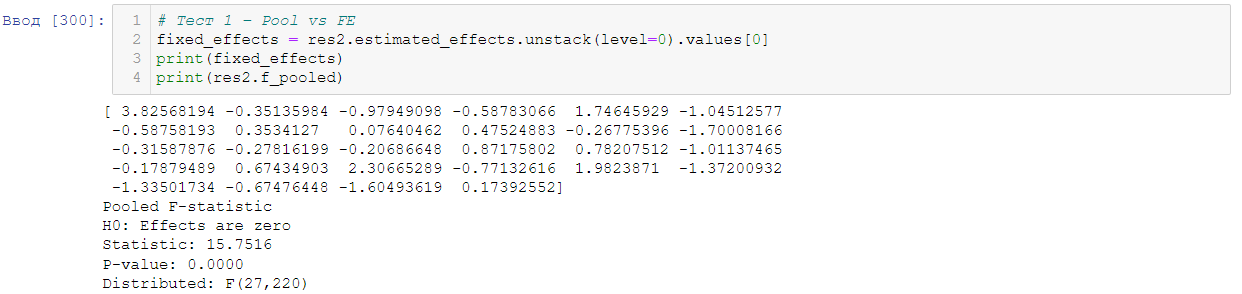
x3 – число прерываний беременности

x4 – коэффициент младенческой смертности 𝑦𝑖𝑡=𝑎𝑖+0.198627𝑥1𝑖𝑡−0.072686𝑥2𝑖𝑡+0.000064𝑥3𝑖𝑡+0.330520𝑥4𝑖𝑡−1.179𝑒−05𝐺𝑅𝑃𝑖𝑡+14.204126+𝑒𝑖𝑡

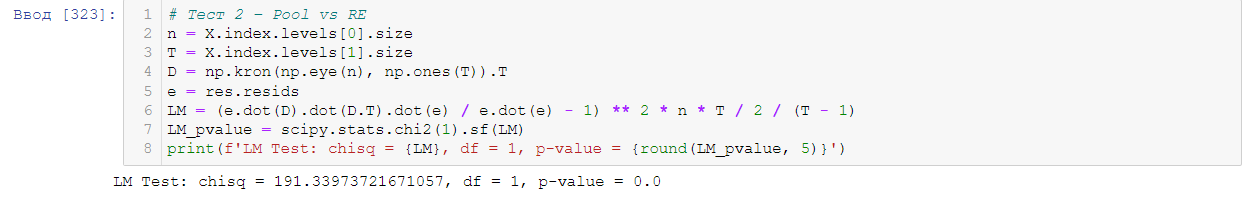
Таблица 3. Меры качества подгонки

|  |  |
| --- | --- |
| R-squared overall | 0,3732 |
| Межгрупповой R-squared(R-squared(between)) | 0,1532 |
| Внутригрупповой R-squared(R-squared(within)) | 0,5377 |
| R-squared | 0,5139 |

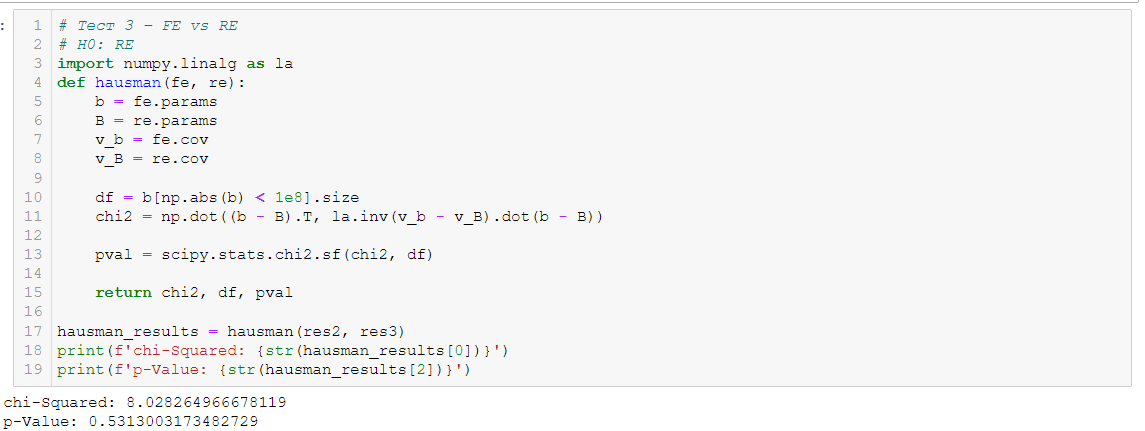
**Тесты**



Так как 𝐹набл> 𝐹крит, 𝑝−𝑣𝑎𝑙𝑢𝑒 <𝛼=0.05, 𝐻0 отклоняется в пользу алтернативной гипотезы, лучшей моделью признается модель с фиксированными эффектами.

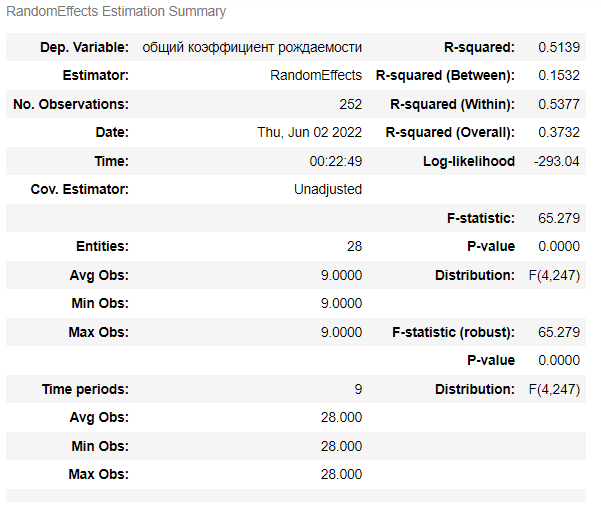


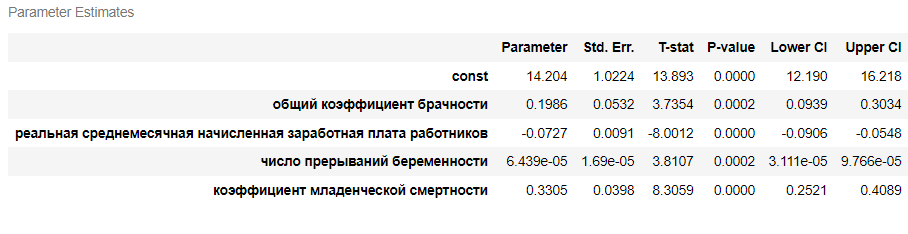
Так как 𝐿𝑀> 𝜒2(1), 𝑝−𝑣𝑎𝑙𝑢𝑒 <𝛼=0.05, 𝐻0 отклоняется в пользу алтернативной гипотезы, лучшей моделью признается модель со случайными эффектами.



Так как 𝑝−𝑣𝑎𝑙𝑢𝑒> 𝛼=0.05, 𝐻0 не отклоняется в пользу алтернативной гипотезы, лучшей моделью признается модель со случайными эффектами.

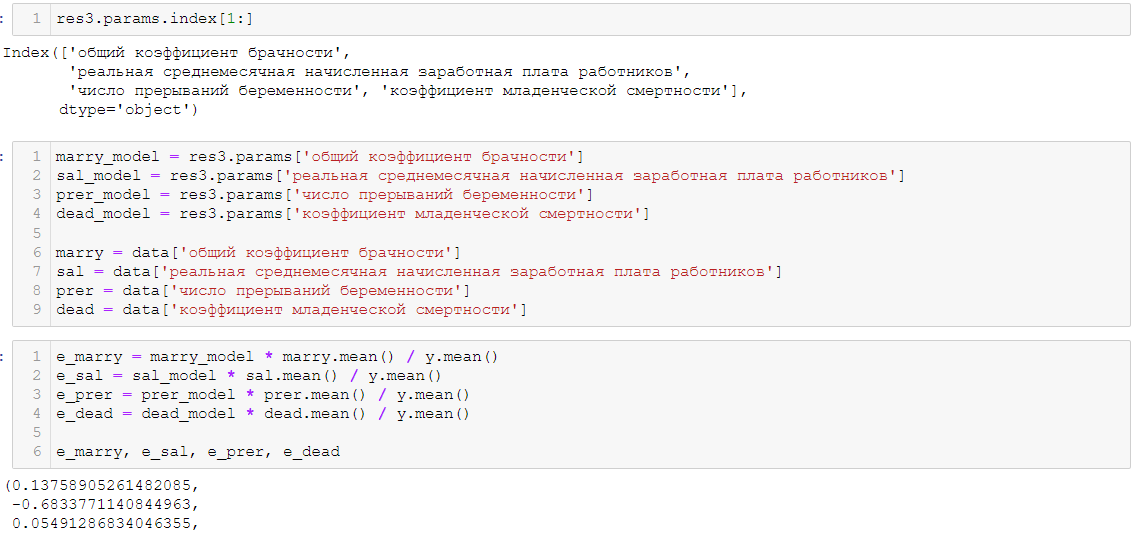
**Финальная модель**



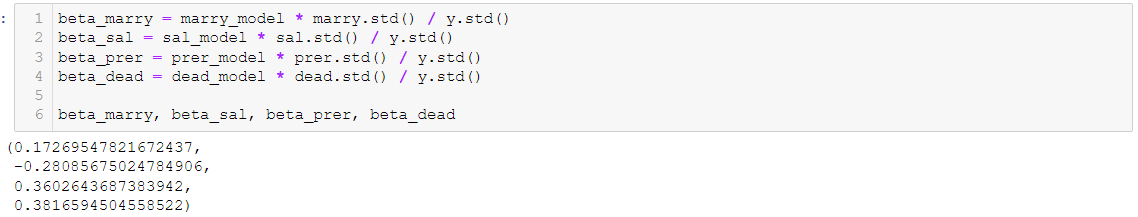


**Расчёт экономических коэффициентов для итоговой модели**

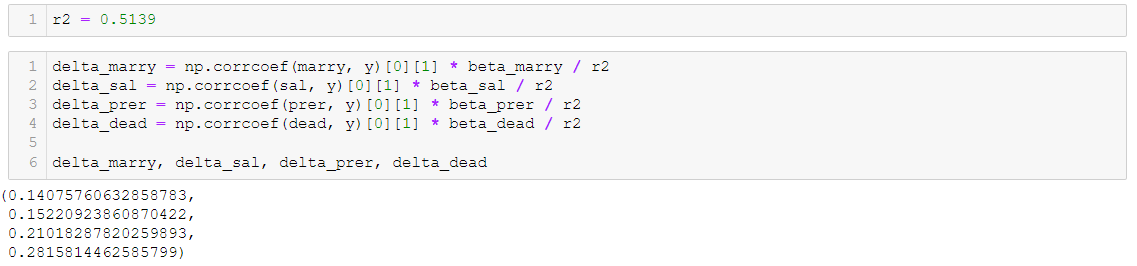
Коэффициент эластичности показывает, на сколько процентов изменится значение исследуемой величины при изменении соответствующего фактора на 1%



Бета-коэффициент показывает, на какую часть своего СКО изменится значение исследуемой переменной при изменении соответствующего фактора на 1 СКО



Дельта-коэффициент показывает среднюю долю влияния соответствующего фактора в совокупном влиянии всех факторов, включенных в модель

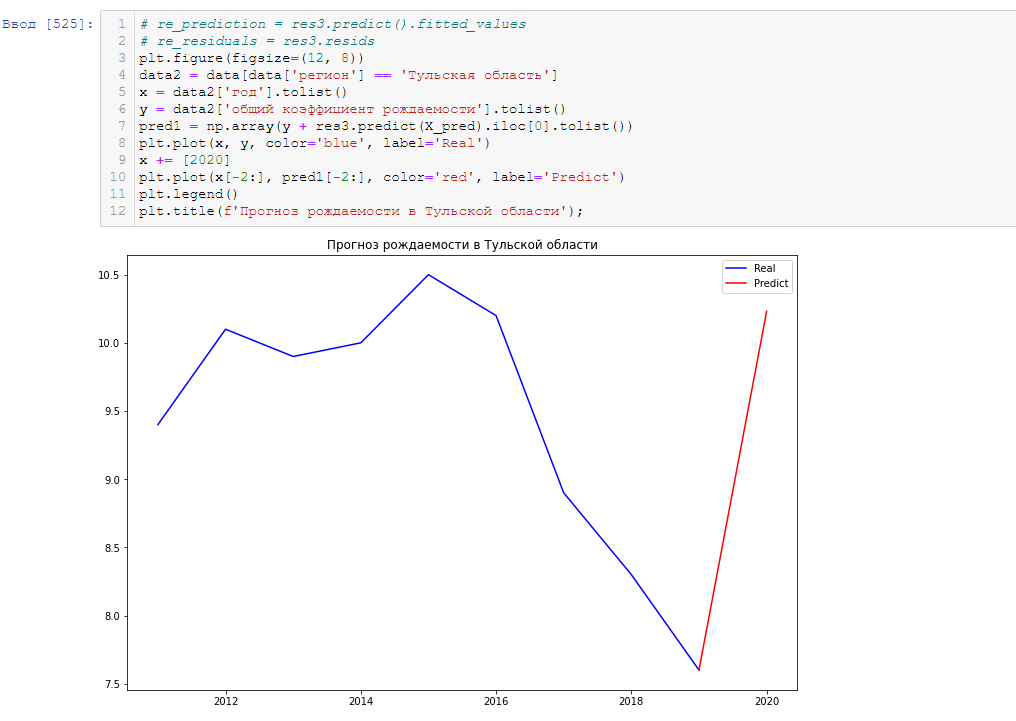


**Предсказание для Тульской области:**

Так как по данным Росстата коэффициент младенческой смертности можно найти только до 2019 года, попробуем аппроксимировать его с помощью линейной регрессии.







1. [Факторы демографического развития России: опыт исследования панельных данных](https://cyberleninka.ru/article/n/faktory-demograficheskogo-razvitiya-rossii-opyt-issledovaniya-panelnyh-dannyh/viewer) [↑](#footnote-ref-1)
2. [Многомерный эконометрический анализ уровня рождаемости в России](https://studbooks.net/2240893/matematika_himiya_fizika/mnogomernyy_ekonometricheskiy_analiz_urovnya_rozhdaemosti_rossii) [↑](#footnote-ref-2)