

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИТМО»**

Отчёт

по лабораторной работе №6 «Бинарная логистическая регрессия»

по дисциплине «**Математические модели исторических процессов**»

Автор: Малаев Степан Геннадьевич

Факультет инфокоммуникационных технологий

Группа: K33422

Преподаватель: Екатерина Ивановна

Санкт-Петербург

2024

Обзор данных

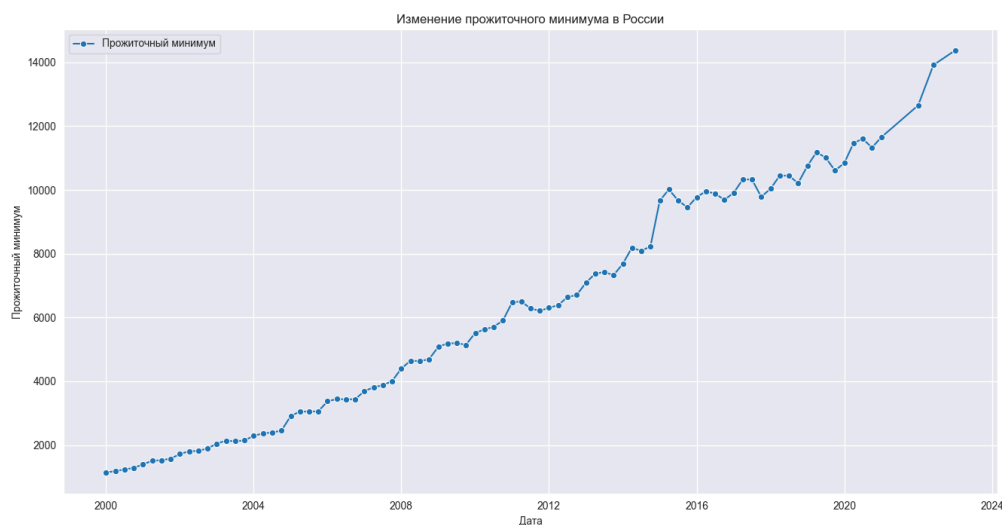
В рамках задания был выбран сценарий влияния на продолжительность жизни выхода на пенсию и минимального потребительского бюджета населения России. В качестве анализируемых данных была собрана статистика за промежуток с 2000 по 2024 год таких метрик как:

- [Средняя продолжительность жизни](#)
- [Средний возраст выхода на пенсию](#)
- [Минимальный прожиточный минимум](#)

Рассмотри данные на графиках:



Средний возраст населения постепенно увеличивался. В 2000 он составлял около 65 лет, а к 2024 вырос примерно до 72-73 лет. Этот рост происходил относительно равномерно, с небольшим замедлением в 2020-2021 годах. Пенсионный возраст в 2000-2018 годах менялся слабо, колеблясь около 59 лет. Однако в 2019 году началось его заметное повышение - с 60 лет в 2019 до 63 лет в 2024 году.



Прожиточный минимум в России вырос с 2000 по 2024 год. В 2000 году он составлял около 1137 рублей, а к 2024 достиг 15453 рублей, увеличившись более чем в 13 раз за 24 года.

Наблюдается корреляция между ростом прожиточного минимума и увеличением среднего возраста населения. Вероятно, повышение уровня жизни способствовало росту продолжительности жизни.

Логистическая регрессия

Для анализа влияния возраста выхода на пенсию и прожиточного минимума на вероятность дожития была сгенерирована синтетическая выборка данных за период с 2000 по 2020 год. Генерация осуществлялась на основе реальных данных о среднем пенсионном возрасте и прожиточном минимуме за каждый год. Возраст выхода на пенсию для каждого человека моделировался случайным образом с отклонением в 1 год от среднего значения. Признак дожития до 75 лет генерировался с вероятностью, линейно возрастающей с 50% в 2000 году до 70% в 2020 году. Такой подход позволил создать правдоподобный набор данных (1000 человек за каждый год) для применения логистической регрессии.

Для оценки совместного влияния возраста выхода на пенсию и прожиточного минимума на вероятность дожития до него была построена модель бинарной логистической регрессии. Зависимой переменной выступал бинарный признак дожития до 75 лет, а независимыми - пенсионный возраст и прожиточный минимум. Модель оценивалась методом максимального правдоподобия с помощью библиотеки [statsmodels](#).

Весь исходный код можно найти в [открытом репозитории проекта](#).

Получившиеся показатели логистической регрессии.

Показатель	Значение
Количество наблюдений	21000
Логарифмическая функция правдоподобия	-13993
Псевдо-R2	0.009293

Коэффициенты модели	Значение
Вероятность дожития (константа)	-0.5247
Возраста выхода на пенсию	0.0090
Прожиточный минимум	6.906e-05

Значимость коэффициентов (на 5% уровне)	Значение
Вероятность дожития (константа)	Незначим
Возраста выхода на пенсию	Незначим
Прожиточный минимум	Значим на 1% уровне p-значение < 0.01

В целом, результаты указывают на то, что более высокий прожиточный минимум ассоциирован с большей вероятностью дожития, в то время как влияние возраста выхода на пенсию не было выявлено. Однако модель обладает низкой объясняющей силой (малый псевдо- R^2), что может быть связано с упрощенным подходом к синтезу данных и неучтенными факторами влияния на продолжительность жизни.

С более качественными данными и выборкой результаты могли получиться намного лучше, но к сожалению, не удалось найти их в открытом доступе.

Вывод

В данной лабораторной работе была исследована взаимосвязь между возрастом выхода на пенсию, прожиточным минимумом и вероятностью дожития до определенного возраста в России с помощью логистической регрессии. Для анализа использовались синтетические данные, сгенерированные на основе реальной статистики за период с 2000 по 2020 год.

Данная работа является лишь иллюстративным примером применения логистической регрессии для анализа влияния социально-экономических факторов на продолжительность жизни. Для получения более надежных выводов необходимо использовать реальные данные на реальной выборке людей, и желательно расширить список учитываемых переменных, например такие как пол, уровень образования, состояние здоровья и тому подобное.