

Рабочее название дипломной работы:

“Прогноз половозрастных уровней участия в рабочей силе”

i. Показатель “уровень участия в рабочей силе”

Уровень участия в рабочей силе – соотношение численности рабочей силы и численности населения в среднем за какой-то период, обычно за год. Несколько лет назад уровни участия в рабочей силе назывались уровнями экономической активности.

Рабочая сила = Занятые в экономике + Безработные (на стр. 3 говорится о том, в каком файле есть точные определения этих понятий – они важны).

Возрастные уровни участия в рабочей силе обычно публикуются для следующих групп (для обоих полов плюс отдельно для мужчин и женщин): 15-19 лет, 20-24 года, ... 55-59 лет, 60-69 лет, 70 лет и старше.

Но, используя микроданные (о них говорится на стр. 2), можно рассчитать уровни участия в рабочей силе для всех однолетних половозрастных групп начиная с 15 лет (кроме самых старших возрастов).

В дипломной работе нужно будет самостоятельно, но только в сторону большей детализации, **определить границы рассматриваемых возрастных групп**. Группы 25-29 лет, 30-34 года, ..., 50-54 года изменять не стоит, но имеет смысл дополнительно разбить группы 15-19 лет, 20-24 года, 55-59 лет и 60-69 лет, в явном виде учитывая школьные и студенческие годы, а также возраст выхода на пенсию.

Папка “01_исходные-данные_ОРС-Росстата”

Данные об участии в рабочей силе разрабатываются Росстатом на основе выборочного “Обследования рабочей силы” (ОРС). Раньше оно называлось “Обследованием населения по проблемам занятости” (ОНПЗ). Проводится обследование ежемесячно, но обновленные микроданные публикуются один раз в год.

В микроданных ОРС Росстат публикует в анонимизированном виде информацию о каждом участвовавшем в обследовании человеке. Дейта-фрейм за каждый год содержит около одного миллиона наблюдений.

В папке “микроданные-2023” собраны микроданные ОРС за 2023 год, список показателей ОРС и методологический комментарий.

С этой страницы Росстата можно скачать микроданные за 2010-2023 годы: https://rosstat.gov.ru/labour_force – Занятость и безработица – Микроданные выборочных обследований рабочей силы.

Что нужно будет сделать в дипломной работе:

- а) изучить эти данные и методологию ОРС;
- б) собрать объединенный дейта-фрейм, содержащий микроданные не только за 2023, но и за весь период 2010-2023 годов, оставив в нем только важные для диплома и одновременно сопоставимые показатели;

в) при необходимости дополнить объединенный дейта-фрейм дополнительными показателями.

Такой объединенный дейта-фрейм – база для дипломных расчетов.

В папке **“агрегированные-данные-2022-2023”** находится росстатовский бюллетень “Итоги выборочного обследования рабочей силы” за 2023 год. В этом бюллетене ежеквартально публикуются агрегированные результаты ОРС. За последней информацией о состоянии рынка труда мы обычно идем сюда, а не в микроданные. Выпуски бюллетеня можно скачать отсюда: <https://rosstat.gov.ru/folder/11110/document/13265>.

В бюллетене есть подробный методологический файл с названием “VVEDENIE3.doc” – его обязательно надо изучить. В частности, чтобы найти точные определения основных ОРС-понятий (занятые, безработные, уровень безработицы т.д.).

Используя таблицы из этого бюллетеня, можно проверить корректность алгоритма, который будет использоваться для агрегирования микроданных ОРС (пример такой проверки содержится в скрипте 01 папки 06).

Папка “02_прогноз-уровней-участия_возможный-подход”

В дипломных расчетах предлагаю отталкиваться от подходов, предложенных в PhD-диссертации Т. Дауда. Детально изучать ее не нужно – например, методику прогнозирования рождаемости можно пропустить. Но обязательно обратите внимание на то, как автор моделирует уровни участия в рабочей силе.

Основное внимание стоит уделить главам 2, 4 и 5 диссертации Т. Дауда. Мы во время обсуждения диплома будем не раз возвращаться к этой работе.

Папка “03_эконометрика_бета-регрессия”

В качестве прогнозно-эконометрической базы диплома предлагаю использовать бета-регрессию.

Бета-регрессия не единственный выбор, логит- и пробит-регрессии могут сначала показаться более очевидными вариантами. О преимуществах бета-регрессии я расскажу во время обсуждения диплома.

В папке две статьи. Изучение бета-регрессии удобнее начать со статьи Дж. Доумы и Дж. Видона.

При изучении бета-регрессии пока стоит пропускать все, что связано с r -values и статистической значимостью. Подробнее об этом я буду рассказывать на лекциях в сентябре, но сейчас можно дополнительно изучить главу 29 учебника Д. Фридмана (я давал ссылку на него в одном из писем).

Папка “04_данные-о-рождаемости-по-возрасту”

В папке два txt-файла с детализированными демографическими данными: возрастные коэффициенты рождаемости и среднегодовая численность женщин. Пример работы с ними - в скрипте 02 папки 06.

Зачем нужна такая статистика? Предлагаю принять рабочую гипотезу, в соответствии с которой вариация уровней участия в рабочей силе определяется (в числе прочих факторов) наличием в семье детей разного возраста. И для проверки этой гипотезы нужны такие данные.

Основной источник детализированной демографической статистики – “Российская база данных по рождаемости и смертности” (РосБРС), она находится здесь: <https://www.nes.ru/demogr-fermort-data?lang=ru>.

В процессе работы над дипломом нужно изучить эту базу данных (txt-файлы в папке 04 скачаны из нее).

Другие полезные демографические источники (ЕМИСС-страницы Росстата):

- численность населения по однолетним возрастным группам:

мужчины – <https://www.fedstat.ru/indicator/31548>;

женщины – <https://www.fedstat.ru/indicator/33459>;

- число родившихся:

<https://www.fedstat.ru/indicator/31606>.

Папка “05_прогнозные-сценарии”

Диплом предполагает построение сценарного прогноза. Сценарий можно разбить на два субсценария: а) демографический; б) экономический.

В папке **“демографический-прогноз-росстата_средний-вариант”** приведены результаты среднего (одного из трех) варианта последнего (от 29 декабря 2023 года) демографического прогноза Росстата. Найти прогноз можно по этой ссылке: <https://rosstat.gov.ru/compendium/document/13285>. Менее детализированные результаты (но для трех вариантов) находятся здесь: <https://rosstat.gov.ru/folder/12781> – Демографический прогноз.

В качестве демографического сценария в дипломе стоит использовать средний вариант демографического прогноза Росстата.

В папке **“среднесрочный-экономический-прогноз_мэр”** приведены результаты последнего среднесрочного прогноза Министерства экономического развития (МЭР). Это самый официальный экономический прогноз в стране. Разрабатывается он обычно для двух сценариев, “базового” и “консервативного”. Обновляется прогноз, как правило, один раз в полгода. Ссылка на прогноз: https://www.economy.gov.ru/material/directions/makroec/prognozy_socialno_ekonomicheskogo_razvitiya/ (документ от 26 апреля 2024 – далее нужно выбрать файл “Приложения”).

В качестве основного экономического сценария в дипломе стоит использовать последний вариант экономического прогноза МЭР.

Папка “06_пример-работы-с-данными_скрипты”

В папке два R-скрипта. В первом – пример агрегирования микроданных ОРС, во втором – пример работы с данными о рождаемости из РосБРС.

Мы их обсудим во время обсуждения диплома.