Федеральное бюджетное государственное образовательное учреждение высшего профессионального образования

«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА

и ГОСУДАРСТВЕННОЙ СЛУЖБЫ

при ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

**ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ**

**ОТДЕЛЕНИЕ ЭКОНОМИКИ**

Кафедра  эконометрики

**ПРОЕКТНАЯ РАБОТА ПО КУРСУ ЭКОНОМЕТРИКА 1**

**«ИССЛЕДОВАНИЕ ГЕНДЕРНОГО РАЗРЫВА В ЗАРАБОТНЫХ ПЛАТАХ».**

Выполнили:

Спирин Владислав Николаевич

Федосеев Роман Сергеевич

Валиахметова Алина Алмазовна

Иванов Пётр Павлович

**МОСКВА, 2024 г.**

**Оглавление**

[1. Данные: 3](#_Toc185471679)

[2. Формулирование гипотез 5](#_Toc185471680)

[3. Построение моделей 5](#_Toc185471681)

[4. Проверка предпосылок 6](#_Toc185471682)

[5. Интерпретация итоговой модели 9](#_Toc185471683)

**Цель исследования:** выяснить, имеется ли гендерный разрыв в заработных платах в РФ

**Задачи исследования:**

1. Изучение теоретической базы
2. Сбор и обработка данных
3. Формулирование гипотезы
4. Построение моделей
5. Проверка предпосылок МНК и ТГМ
6. Интерпретация результатов итоговой модели

# 1. Данные:

Данные берутся за 2012 и 2021 года по двум регионам, Ростовская область и Санкт-Петербург.

*Регрессоры:*

* children – количество детей
* Pension – получаете ли вы пенсию (Бинарная)
* Main\_activity - Какой ответ лучше всего описывает Ваше основное занятие в настоящее время?
* Take\_loan – брали ли вы кредит за последние 12 месяцев (бинарная)  
  Sex – Пол (бинарная)
* Age – возраст
* Diplom – законченное образование (какой диплом у человека)
* Family – Состоит ли в браке
* Proff\_group – профессиональная группа (Квалифицированный или неквалифицированный работник – колонка была перекодирована по следующему принципу: чиновники, специалисты высшего уровня и квалифицированные рабочие это в одну группу, все остальные в другую) (бинарная)
* Is\_ceo – есть ли у человека подчиненные (бинарная)
* foreign\_language– владеет ли опрашиваемый каким нибудь языком помимо стран бывшего СССР (бинарная)
* Add\_ins – имеет ли человек дополнительное страхование жизни (бинарная)
* job\_satisfaction – Удовлетворен ли человек своей работой (бинарная)

*Зависимая переменная:*

**income\_30d** - Сколько денег в течение последних 30 дней Вы получили по основному месту работы после вычета налогов?

Статистическое описание данных:

* Посмотрим описательные статистики для 2012 года обработанных данных. Укажем количество данных для каждого региона:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Ростов: 155 наблюдений

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Санкт-Петербург 182 наблюдения

* Посмотрим описательные статистики для 2021 года обработанных данных. Укажем количество данных для каждого региона:

Изображение выглядит как текст, снимок экрана, черно-белый, Шрифт

Автоматически созданное описание

Ростов: 168 наблюдений

Изображение выглядит как текст, снимок экрана, черно-белый, число

Автоматически созданное описание

Санкт-Петербург: 150 наблюдений

# 2. Формулирование гипотез

H0 : β1 = 0 – отсутствие гендерного разрыва в регионах

H1 : β1 ≠ 0 – наличие гендерного разрыва в регионах РФ по заработным платам

# 3. Построение моделей

Были построены 4 модели для каждого региона и года отдельно. Модели были оценены методом МНК. Целевая переменная – логарифм заработных плат, регрессор – бинарная переменная пола. Остальные переменные были объявлены контрольными, для уменьшения смещения оценок и повышения точности оценки. Рассмотрим более подробно построение модели для Ростовской области 2012 года.

Ln(income\_30d) = β0 + β1 ⋅ sex + β2 ⋅ pension + β3 ⋅ main\_activity + β4 ⋅ take\_loan + β5 ⋅ children + β6 ⋅ age + β7 ⋅ diplom + β8 ⋅ family + β9 ⋅ proff\_group + β10 ⋅ is\_ceo + β11 ⋅ job\_satisfaction + β12 ⋅ age2 + ui

Изображение выглядит как текст, число, Шрифт, чек

Автоматически созданное описание

Модель для Ростовской области за 2021 год:

ln(income\_30d) = β0 + β1 ⋅ sex + β2 ⋅ age+ β3 ⋅ age^2 + β4 ⋅ children + β5 ⋅ diplom + β6 ⋅ family + β7 ⋅ is\_ceo + β8 ⋅ job\_satisfaction + β9 ⋅ main\_activity + β10 ⋅ proff\_group + β11 ⋅ work\_day\_hours +ui

Модель для Санкт-Петербурга за 2012 год:

ln(income\_30d) = β0 + β1 ⋅ sex + β2 ⋅ age + β3 ⋅ age^2+ β4 ⋅ children + β5 ⋅ diplom + β6 ⋅ family + β7 ⋅ is\_ceo + β8 ⋅ job\_satisfaction + β9 ⋅ main\_activity + β10 ⋅ pension + β11 ⋅ proff\_group + β12 ⋅ take\_loan + ui

Модель для Санкт-Петербурга за 2021 год:

ln(income\_30d) = β0 + β1 ⋅ sex + β2 ⋅ add\_ins + β3 ⋅ age+ β4 ⋅ age^2 + β5 ⋅ diplom + β6 ⋅ family + β7 ⋅ for\_ language + β8 ⋅ is\_ceo + β9 ⋅ job\_satisfaction + β10 ⋅ main\_activity + β11 ⋅ proff\_group + β12 ⋅ work\_exp +ui

# 4. Проверка предпосылок

Тесты также будем предоставлять для данной итоговой модели (Ростовская область, 2012 год), чтобы не смешивать все в кучу. Предпосылки ТГМ выполняются для каждой из итоговых моделей, интерпретация приведена в следующем пункте.

*VIF-тест*

Для проверки мультиколлинеарности мы использовали VIF-тест, значение меньше 10 указывает на отсутствие высокой мультиколлинеарности, наблюдаем что тест показывает, что у нас ее нет

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

*тест Бройша-Пагана и тест Уайта*

Проверка гомоскедастичности – провели тест Бройша-Пагана и тест Уайта:



Лагранжева мультипликативная статистика (LM Statistic) используется для проверки гипотезы о гомоскедастичности.

Если значение p-value меньше уровня значимости (например, 0.05), то гипотеза о гомоскедастичности отвергается, и можно предположить наличие гетероскедастичности.

F-Statistic: Статистика F-теста. Используется для проверки гомоскедастичности.

Если F-Test p-value меньше уровня значимости (например, 0.05), то гипотеза о гомоскедастичности отвергается.

Test Statistic: Статистика теста Уайта.



Если p-value меньше уровня значимости (например, 0.05), то гипотеза о гомоскедастичности отвергается, и можно предположить наличие гетероскедастичности.

F-Statistic: Статистика F-теста.

F-Statistic p-value: Значение p-value для F-Statistic. Если это значение меньше уровня значимости (например, 0.05), то гипотеза о гомоскедастичности отвергается.

*Тест Шапиро-Уилка*

**

Графики сверху:

Что показывают: Эти графики сравнивает квантильное распределение остатков модели с квантильным распределением нормального распределения.

Как интерпретировать: Если точки на графиках лежат примерно на прямой линии, это указывает на то, что остатки модели следуют нормальному распределению. Отклонения от прямой линии указывают на отклонения от нормальности.

Гистограмма остатков отображает распределение остатков модели.

Как интерпретировать: позволяет визуально оценить, насколько распределение остатков похоже на нормальное распределение.

*Изображение выглядит как диаграмма, График, линия, текст

Автоматически созданное описание*

*Изображение выглядит как диаграмма, текст, линия, График

Автоматически созданное описание*

Применяется для проверки нормальности распределения остатков.

Подходит для малых и средних выборок данных.

если p-value > 0.05, то данные нормально распределены

# 5. Интерпретация итоговой модели

1. Основные результаты:

• Итоговая модель объясняет 60% вариации доходов, что свидетельствует о её высокой объясняющей способности.

• Наиболее значимые факторы, влияющие на доход, включают:

* Пол
* Возраст
* Основная деятельность
* Удовлетворённость работой

2. Возможные пропущенные переменные, которые могут повлиять на результаты:

* Доступ к гибким графикам работы: женщины могут выбирать работу с меньшим доходом, но более удобными условиями.
* Часы работы: если мужчины работают больше часов, это может объяснять часть разрыва.
* Социальные нормы и ожидания: они могут влиять на карьерные предпочтения. Например, при ожидании повышения в карьере человек получает больше мотивации работать, что увеличивает его производительность и, следовательно, влияет на уровень зп.

• Если пропущенные переменные коррелируют с регрессорами (например, полом или возрастом) и зависимой переменной, это приведёт к смещению оценок.

В Ростовской области в 2012 и 2021 году замечен гендерный разрыв в заработных платах. В Санкт-Петербурге гендерный разрыв наблюдается лишь в 2012 году.

На основании итоговой модели мы можем заключить, что в 2012 году в Ростовской области женщины зарабатывали в среднем на 60,3% меньше мужчин, что свидетельствует об очень сильном гендерном разрыве в заработных платах. Однако в 2021 году этот коэффициент уменьшился, что привело к разрыву в заработных платах в 34,4%. Возможными причинами может быть довольно развитая промышленность и транспортная сеть в области, где уровень заработных плат выше, но так как работа связана с физическим трудом и опасностями, отдается предпочтение мужскому полу. Для Санкт-Петербурга в 2012 году женщины зарабатывали в среднем на 7,1% меньше мужчин, что является гораздо меньшим разрывом, чем по Ростовской области и свидетельствует о несильной дискриминации по половому признаку. Однако в 2021 году, коэффициент оказывается незначимым, что свидетельствует о равенстве в уровне заработных платах для мужчин и женщин.