# Аналитическая записка

## Тема

Построение и оценка эконометрической модели зависимости среднедушевых денежных доходов населения от средней номинальной заработной платы

## Цель анализа

Исследовать взаимосвязь между средней номинальной заработной платой (WAGE\_Q) и среднедушевыми доходами населения (HHI\_Q), оценить спецификацию модели, проверить выполненные предпосылки и построить прогнозное значение с доверительным интервалом.

## Задание 1: Построение спецификации эконометрической модели

Постановка задачи: Мы выбрали среднедушевые денежные доходы населения (HHI\_Q) в качестве зависимой переменной, а среднюю номинальную заработную плату (WAGE\_Q) — в качестве независимой переменной. Построена диаграмма рассеяния и рассчитан коэффициент корреляции.  
  
Результаты:  
- Коэффициент корреляции между WAGE\_Q и HHI\_Q составил 0.9640, что указывает на сильную положительную связь между переменными.  
- p-значение: 3.84e-21, что значительно ниже уровня значимости 0.05, подтверждая статистическую значимость корреляции. Это указывает на высокую вероятность того, что связь между переменными не случайна.

## Задание 2: Оценивание качества спецификации модели

Построение модели: Была построена линейная регрессионная модель, описывающая зависимость HHI\_Q от WAGE\_Q. Результаты оценки параметров модели:  
  
- Константа: -294.03, p-значение = 0.800, что указывает на отсутствие значимости константы на уровне 0.05.  
- Коэффициент при WAGE\_Q: 0.8816, p-значение < 0.0001. Это значение показывает, что средняя номинальная заработная плата оказывает статистически значимое влияние на доходы населения.  
- Коэффициент детерминации (R^2): 0.929, что означает, что модель объясняет 92.9% изменчивости доходов населения на основе заработной платы.  
  
Средняя относительная ошибка аппроксимации составила 5.44%, что подтверждает высокую точность модели в описании данных.  
  
Вывод: Модель продемонстрировала высокую объяснительную способность, так как большая часть изменчивости доходов населения объясняется изменениями в средней номинальной заработной плате.

## Задание 3: Проверка остатков на нормальное распределение

Методология: Для проверки нормальности распределения остатков использован тест Шапиро-Уилка. Этот тест выбран, так как он хорошо подходит для небольших и средних выборок и чувствителен к отклонениям от нормальности.

Для проверки нормальности распределения остатков выбран **тест Шапиро-Уилка**. Основные причины его использования заключаются в следующем:

1. **Подходит для малых и средних выборок**: Тест Шапиро-Уилка считается одним из наиболее мощных тестов на нормальность для небольших и средних выборок, особенно когда количество данных не очень велико. Он оптимизирован для этих случаев, что делает его более надежным в сравнении с другими тестами.
2. **Высокая чувствительность к отклонениям от нормальности**: Шапиро-Уилк имеет высокую чувствительность к различным типам отклонений от нормального распределения. Это важно при анализе остатков, так как даже небольшие отклонения от нормальности могут повлиять на выводы о корректности модели.
3. **Частое применение в эконометрическом анализе**: Этот тест является стандартом для проверки нормальности остатков в эконометрии и статистике. Его результаты легко интерпретировать, и он широко поддерживается в аналитических пакетах, таких как Python и R.

Результаты:  
- Статистика теста Шапиро-Уилка: 0.9717  
- p-значение: 0.4748  
  
Так как p-значение больше 0.05, у нас нет оснований отвергать гипотезу о нормальности распределения остатков. Это означает, что остатки модели можно считать нормально распределёнными.  
  
Вывод: Остатки нормально распределены, что подтверждает выполнение предпосылки нормальности для модели.

## Задание 4: Проверка предпосылки теоремы Гаусса-Маркова о гомоскедастичности

Методология: Для проверки гомоскедастичности проведён тест Бреуша-Пагана, который проверяет постоянство дисперсии остатков. Этот тест выбран, так как он эффективно выявляет гетероскедастичность в линейных моделях.

Для проверки гомоскедастичности выбран **тест Бреуша-Пагана**, потому что:

1. **Прямо тестирует гетероскедастичность** — проверяет, меняется ли дисперсия остатков в зависимости от независимых переменных.
2. **Подходит для линейных моделей** — часто используется для оценки соответствия остатков предпосылке постоянной дисперсии.
3. **Удобен для интерпретации** — результат дает статистику и pp-значение, что позволяет легко определить наличие или отсутствие гетероскедастичности.

Тест Голдфелда-Квандта не использован, так как он менее универсален и предполагает влияние только одной переменной на гетероскедастичность.

Результаты:  
- Статистика Бреуша-Пагана: 5.93  
- p-значение: 0.0149  
  
Так как p-значение меньше 0.05, мы отвергаем нулевую гипотезу о гомоскедастичности и заключаем, что в модели наблюдается гетероскедастичность.  
  
Вывод: Наличие гетероскедастичности означает, что дисперсия ошибок меняется в зависимости от значений независимой переменной. Это может повлиять на точность стандартных ошибок, и для корректных выводов могут потребоваться робастные стандартные ошибки.

## Задание 5: Построение прогнозного доверительного интервала

Цель: Рассчитать прогнозное значение среднедушевых доходов населения при уровне заработной платы, равном 105% от её среднего значения, а также построить доверительный интервал для этого прогноза.  
  
Результаты:  
- Среднее значение WAGE\_Q: примерно 26771.89  
- 105% от среднего значения WAGE\_Q: 28110.48  
- Точечное прогнозное значение HHI\_Q при WAGE\_Q = 28110.48 составляет 24489.01.  
- Доверительный интервал (95%) для прогнозного значения: [23900.74, 25077.29].  
  
Вывод: Прогнозируемое значение среднедушевых доходов населения при увеличении средней заработной платы на 5% от среднего уровня составляет 24489.01. Доверительный интервал на уровне значимости 0.05 показывает, что при данном уровне заработной платы доходы населения с вероятностью 95% будут находиться в диапазоне от 23900.74 до 25077.29.

## Задание 6:

1. **Коэффициент корреляции** между средней заработной платой (WAGE\_C\_Q) и доходами населения (HHI\_Q) составляет 0.9640, что указывает на сильную положительную связь. pp-значение (3.84e-21) подтверждает значимость корреляции.
2. **Модель линейной регрессии**:
   * **Константа**: -294.03 (не значима, p-значение = 0.800).
   * **Коэффициент при WAGE\_C\_Q**: 0.8816 (значим, pp-значение < 0.0001).
   * **R2R2**: 0.929, что указывает на высокую объяснительную способность модели (92.9% изменчивости доходов объясняется зарплатой).
3. **Средняя относительная ошибка аппроксимации**: 5.44%, что подтверждает точность модели.

**Заключение**: Средняя заработная плата оказывает значительное влияние на доходы населения, и модель может быть полезна для прогнозов и анализа экономической политики.

## Общий вывод

В ходе анализа была построена эконометрическая модель, описывающая зависимость среднедушевых доходов населения от средней номинальной заработной платы. Модель показала высокую объяснительную способность, а коэффициент при заработной плате оказался значимым. Проверки на нормальность остатков показали соответствие предпосылке нормальности, однако тест Бреуша-Пагана выявил гетероскедастичность остатков, что может требовать дальнейшего исследования.  
  
Полученные прогнозные значения и доверительный интервал могут быть полезны для оценки возможного уровня доходов населения при изменении заработной платы, что может послужить основой для принятия решений в экономической политике и планировании.