



**КОМАНДА №7 "UNO"**

МУХАМЕТЗЯНОВ БУЛАТ

ТЕРЕНТЬЕВ КОНСТАНТИН

ХУЗИЯХМЕТОВ ТИМУР

ДАЩЕНКО ДМИТРИЙ

ЗУЕВ АЛЬБЕРТ

ХИСАМИЕВ ИЛЬНАЗ

## Датасет: RLMS HSE от 2016 г.

- Регион
- Тип населённого пункта
- Возраст
- Пол (0/1)
- Отрасль
- Логарифм заработной платы
- Гос. компания (0/1)
- Имеет доступ к интернету (0/1)
- Количество детей
- Уровень образования
- + дамми-переменные

**3871** записей

**31** отрасль

**38** регионов

**75** дамми-переменных

Датасет собран на основе опроса населения, который проводился ВШЭ в 2016 г.

# Исследовательский вопрос



ОТ ЧЕГО ЗАВИСИТ  
ЗАРАБОТНАЯ ПЛАТА  
(ДОХОД)?

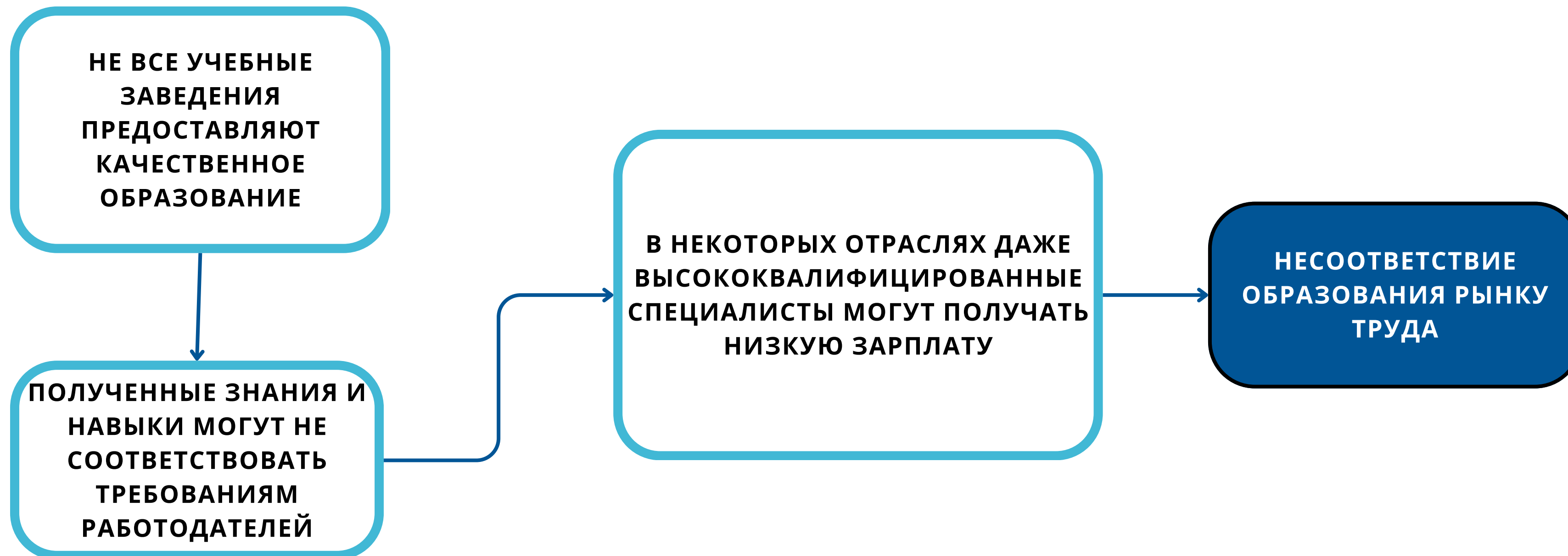
# Гипотеза

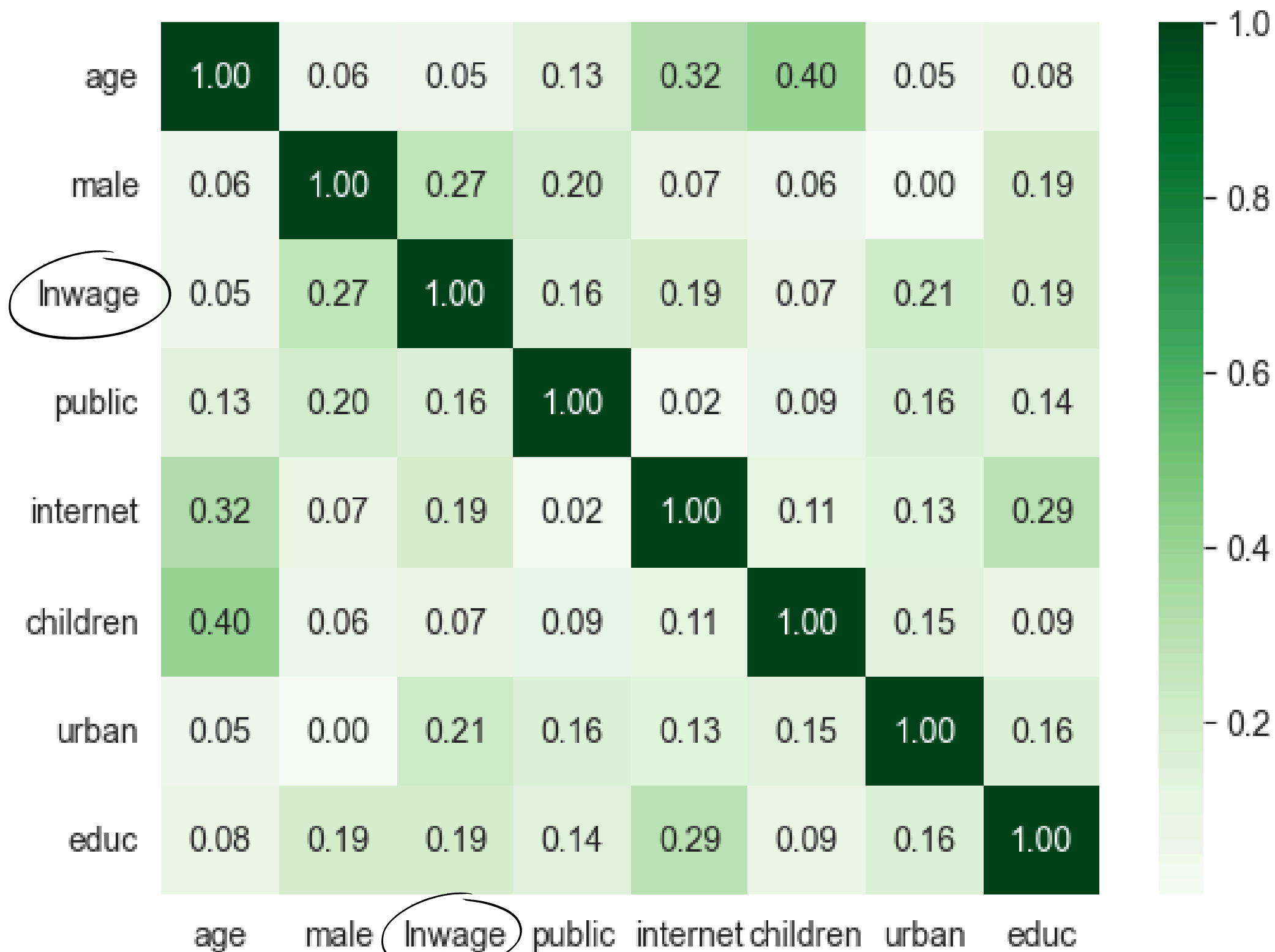
## Доход зависит от уровня образования

**H0:** Уровень образования незначимо влияет на доход

**H1:** Уровень образования значимо влияет на доход

# Механизм



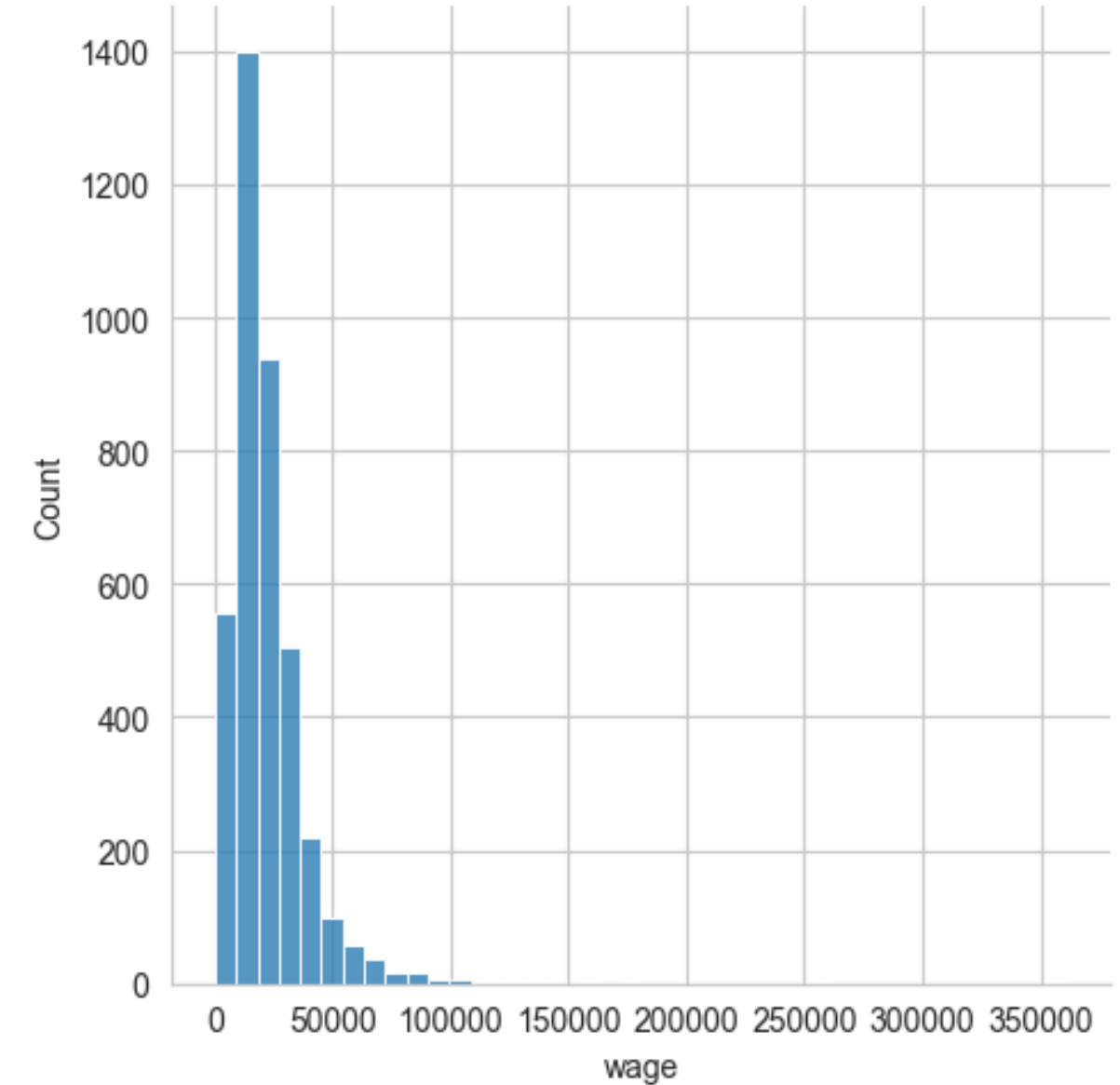
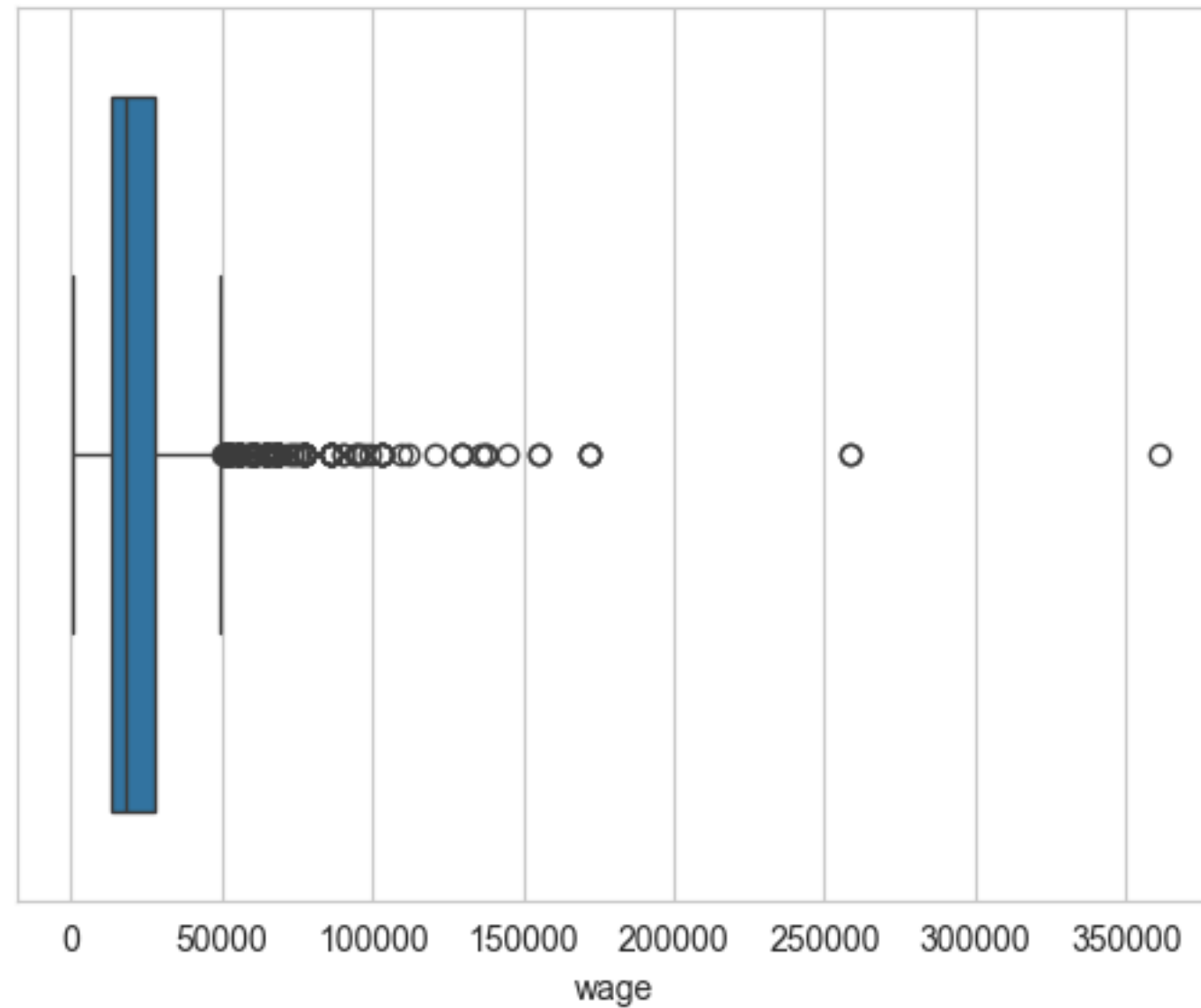


**Отсутствие  
зависимостей  
с доходом**

# Доходы

Добавили  
параметр `wage`,  
исходящий из  
`lnwage`

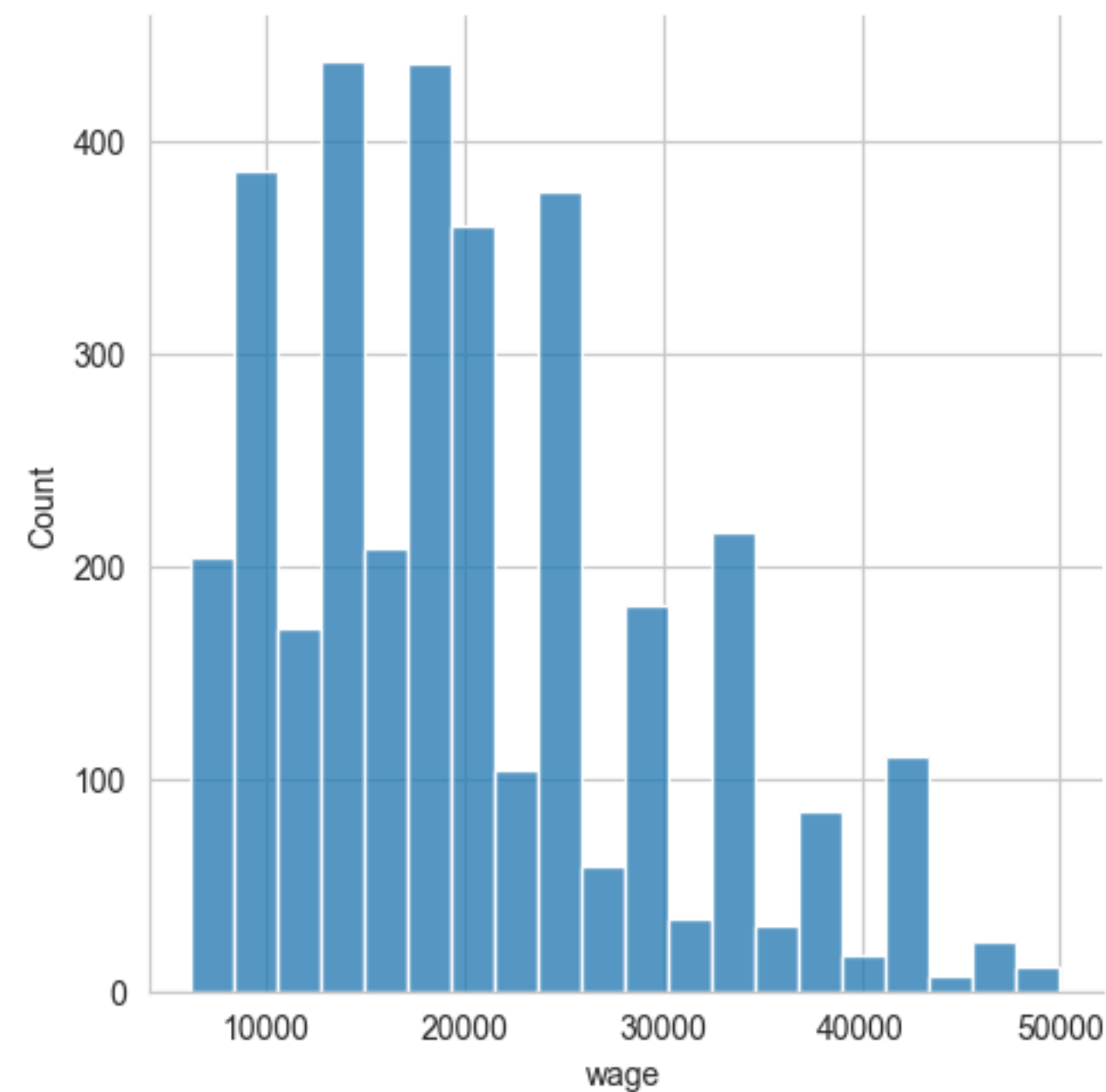
$$\text{wage} = \exp(\text{lnwage})$$



# Работа с выбросами

**Удалили записи  
по 5 и 95 перцентилю  
ДОХОДОВ**

**Дубликаты не  
обнаружены**



Доходы после фильтрации





# Коэффициент корреляции Спирмена

**x: educ** (уровень образования)  
**y: lnwage** (логарифм дохода)

Корреляция  $\approx 0.178$

P-value  $\approx 0$

**P-value < 0.05;**  
**данные не случайны**

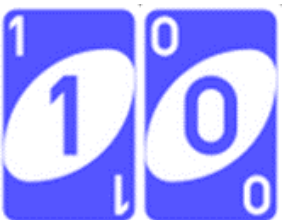
# Линейная регрессия дохода от уровня образования

Доход: **wage** (количественная)

Образование: **educ** (категориальная),  
разбили на 4 dummy-переменные

0	9 классов
1	11 классов
2	ПТУ
3	Высшее

$$\begin{aligned} \mathbf{wage} = & \beta_{\text{const}} + \\ & + \beta_0 * \mathbf{educ\_0} + \\ & + \beta_1 * \mathbf{educ\_1} + \\ & + \beta_2 * \mathbf{educ\_2} + \\ & + \beta_3 * \mathbf{educ\_3} \end{aligned}$$



# Линейная регрессия дохода от уровня образования

## Коэффициенты

$\beta_{const}$	2.08e+12
$\beta_0$	2.08e-12
$\beta_1$	2.08e-12
$\beta_2$	2.08e-12
$\beta_3$	2.08e-12

## Результаты

$$R^2 \approx -0.007$$

$$P\text{-value} \approx 0$$

**$P\text{-value} < 0.05;$   
данные не случайны**



## Основные результаты

### Корреляция

$$|r| < 0.25;$$

связь очень слабая или  
отсутствует

### Линейная регрессия

$$|R^2| < 0.25;$$

связь очень слабая или  
отсутствует

Принимаем  $H_0$ ; отвергаем  $H_1$ .

**Гипотеза не подтвердилась**

## Устойчивость модели

Разделили датасет по федеральным округам

ФО	R <sup>2</sup>	P-value	Вывод
УФО-СЗФО	-0.015	0.59	Нельзя сделать вывод, p-value > 0.05
СКФО-ЮФО	0.042	≈0	Принимаем H <sub>0</sub> (низкий R <sup>2</sup> )
ПФО	0.054	≈0	Принимаем H <sub>0</sub> (низкий R <sup>2</sup> )
ЦФО	0.017	≈0	Принимаем H <sub>0</sub> (низкий R <sup>2</sup> )
ДВФО-СФО	0.080	≈0	Принимаем H <sub>0</sub> (низкий R <sup>2</sup> )

## Перспективы



Сравнение влияния уровня образования на доход в разных регионах



Провести повторное исследование через 10 лет используя актуальные данные и сравнить результаты с текущими

## Ограничения



В исследовании не были учтены другие факторы такие как навыки, опыт работы, социальное происхождение, место проживания, дискриминация и т. д.



Данные были собраны 8 лет назад, поэтому они могут быть неактуальными

# Применение

1

Оценка эффективности экономических моделей на основе данной

2

Натолкнуть на изменение системы образование

# Policy Implication

Реформа системы образования

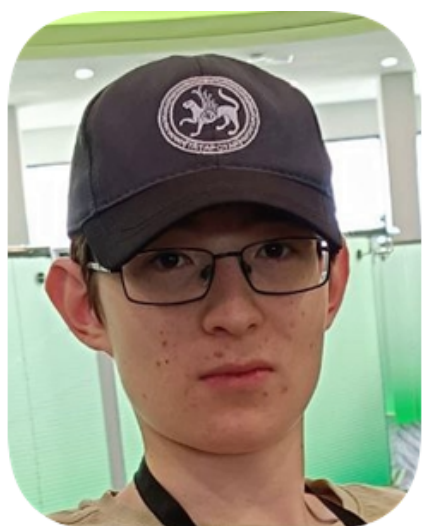
Например, адаптация учебных планов под требования бизнеса

# Альтернативный механизм





# НАША КОМАНДА



Мухаметзянов Булат  
тимлид, программист



Дашенко Дмитрий  
генератор идей



Тереньтев Константин  
программист



Хузияхметов Тимур  
программист



Зуев Альберт  
мат.модель



Хисамиев Ильназ  
коммуникатор



# Приложение

# Доля федеральных округов

