



Selling Pandas

1. Иванов Герман
2. Иванова Полина
3. Михайлов Андрей
4. Кириченко Дмитрий
5. Спирихин Андрей
6. Кром Илья

Алло, меня слышно? А запись будет?



Структура данных

Информация

Лонгитюдное индивидуальное
обследование РМЭЗ НИУ ВШЭ

Расширенный датасет¹

3871 респондент, 110 столбцов²

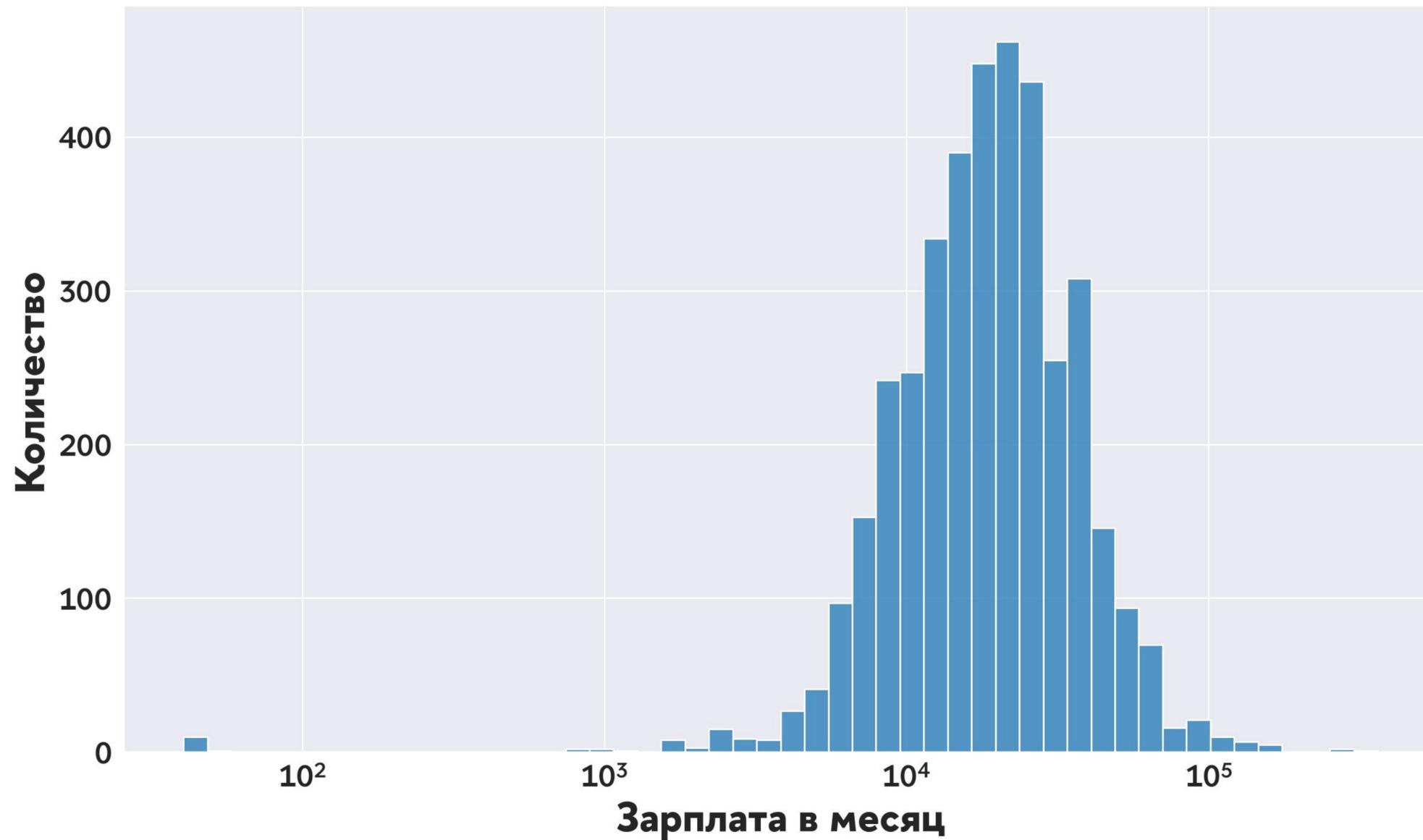
Данные за 2017 год





Предварительный анализ

Распределение зарплат респондентов



Среднее: 22589.9

Медиана: 18074.9

Стандартное отклонение: 17844.7





Тест Шапиро-Уилка

F0

Генеральная совокупность обладает нормальным распределением

F1

Генеральная совокупность не обладает нормальным распределением

Результат

Показатель	Значение	F0
p-value	3.78×10^{-44}	✗

Вывод

Тест Шапиро-Уилка на нормальное распределение не пройден

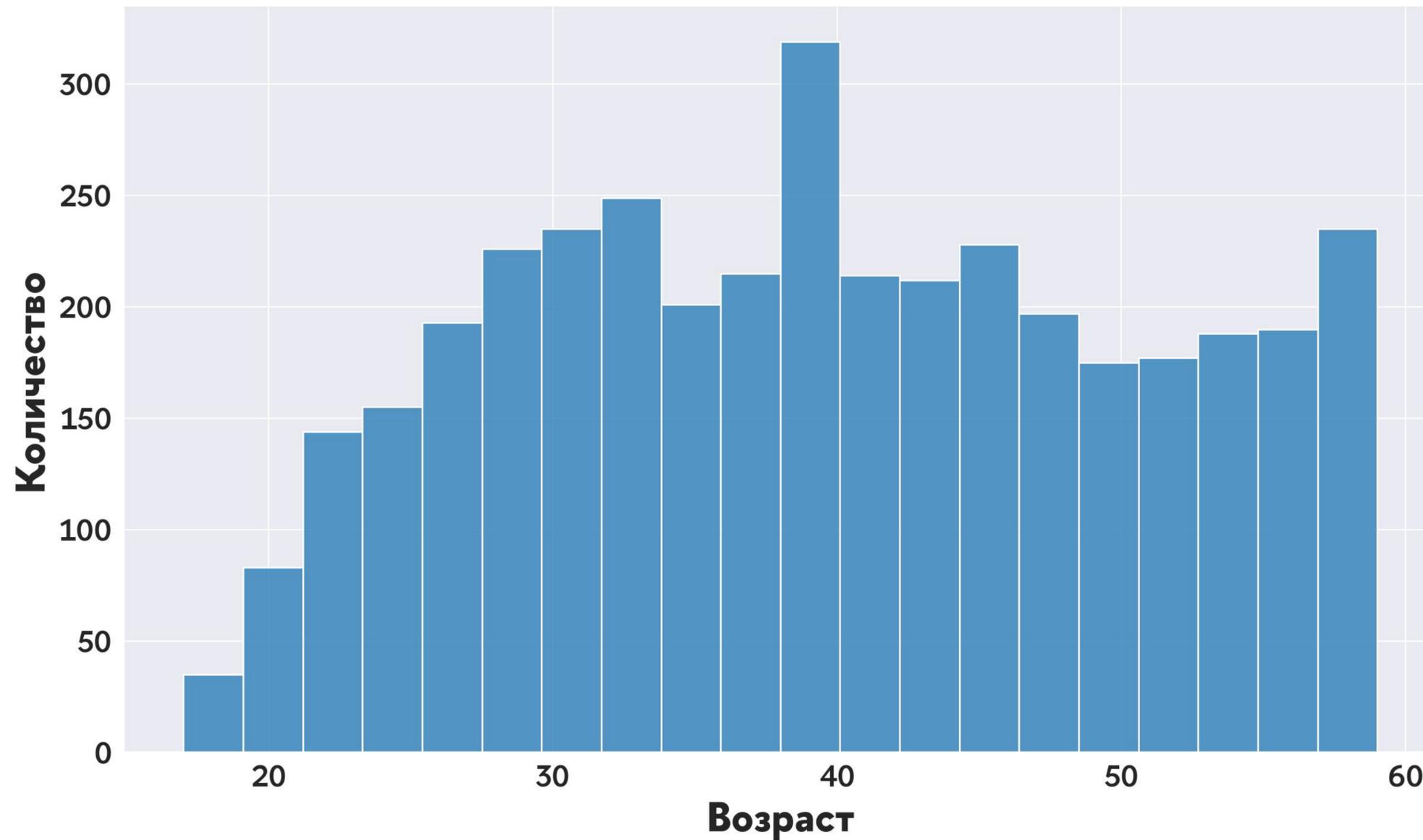


Мы будем использовать непараметрический U-критерий Манна-Уитни



Предварительный анализ

Распределение возраста респондентов



Среднее: 39.6

Медиана: 39.0

Стандартное
отклонение: 10.8





Матрица корреляций

Логарифм зарплаты	1	-0.16	0.19	0.27	-0.05	-0.072
Гос-венная ли компания	-0.16	1	0.14	-0.2	0.13	0.094
Уровень образования	0.19	0.14	1	-0.19	-0.077	-0.09
Мужчина ли человек	0.27	-0.2	-0.19	1	-0.062	-0.058
Возраст	-0.05	0.13	-0.077	-0.062	1	0.4
Кол-во детей	-0.072	0.094	-0.09	-0.058	0.4	1
	Логарифм зарплаты	Гос-венная ли компания	Уровень образования	Мужчина ли человек	Возраст	Кол-во детей





Обработка данных

Замена заголовков

psu



region

educ

education_level

Расширение датасета

Merge колонок





Очистка от выбросов

Очистка

По заработной плате

Отбрасываем по 2,5% сверху и снизу

3871 строка → 3679 строк

Оставляем в итоговой выборке тех, кто обучался в вузе (`education_level = 3`)

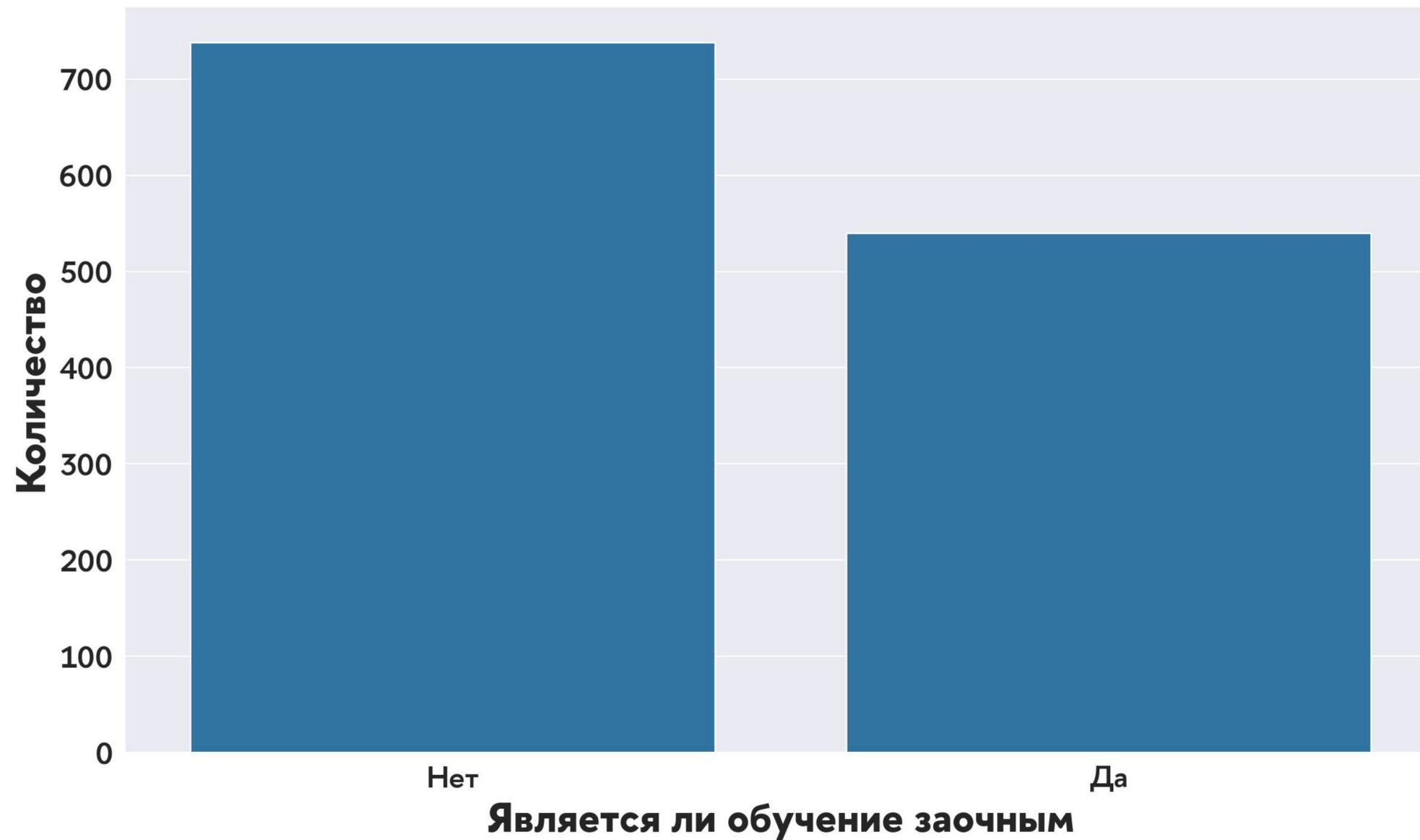
3679 строк → 1278 строк





Предварительный анализ

Распределение формата обучения в ВУЗе



Очная форма: 799

Заочная форма: 583





Исследовательский вопрос и гипотеза

Исследовательский вопрос

Как на **заработную плату** влияет **форма обучения** при получении высшего образования?

Гипотеза

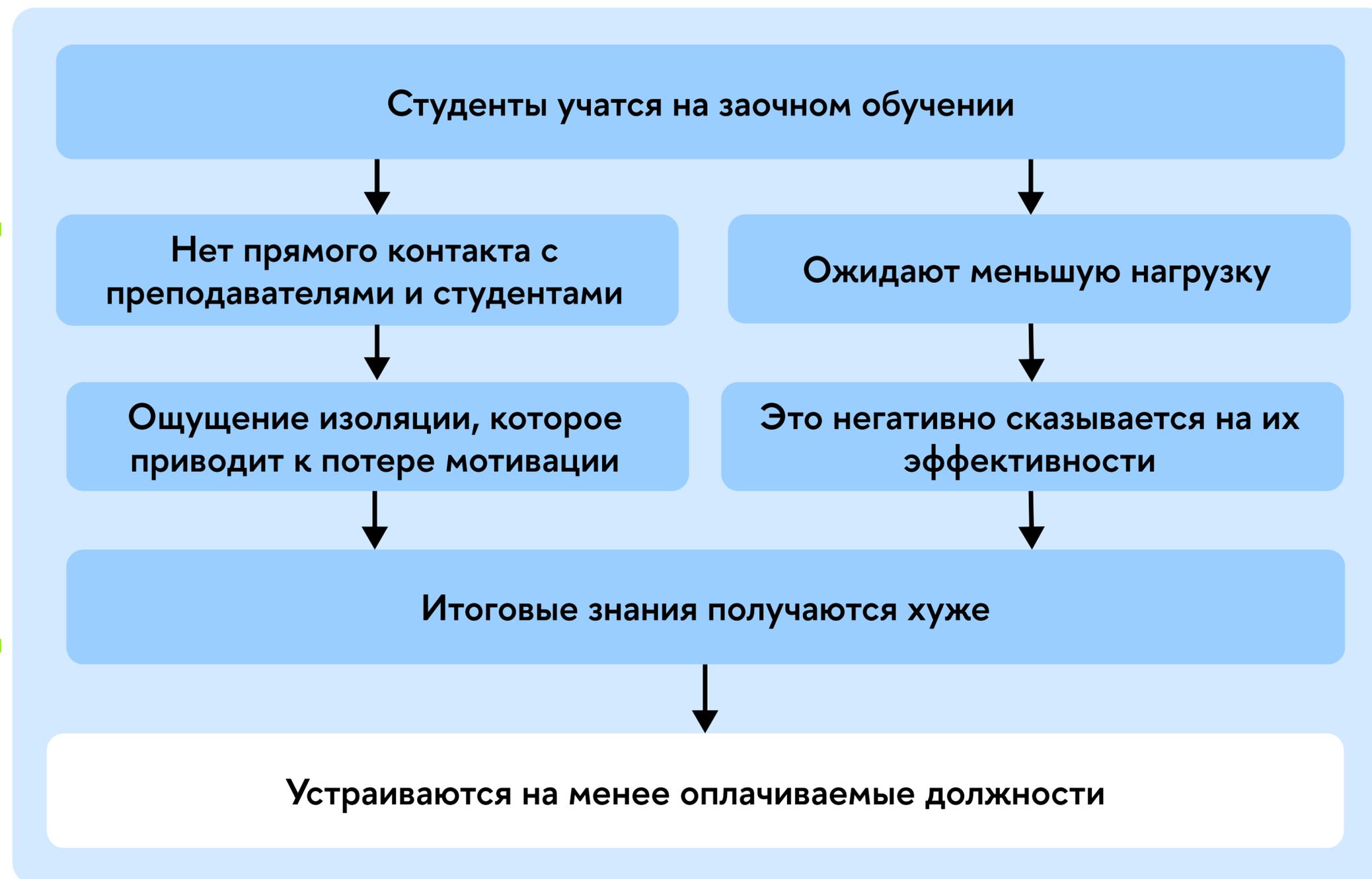
При **заочной форме** обучения на программах высшего образования **заработная плата ниже**, чем при очной форме обучения





Механизмы

**Первый
механизм**



**Второй
механизм**



Математическая модель

Итоговый объем выборки

1278

Уровень стат. значимости

5%

H0

Зарплаты людей, получивших образование на очной и на заочной формах образования **равны**

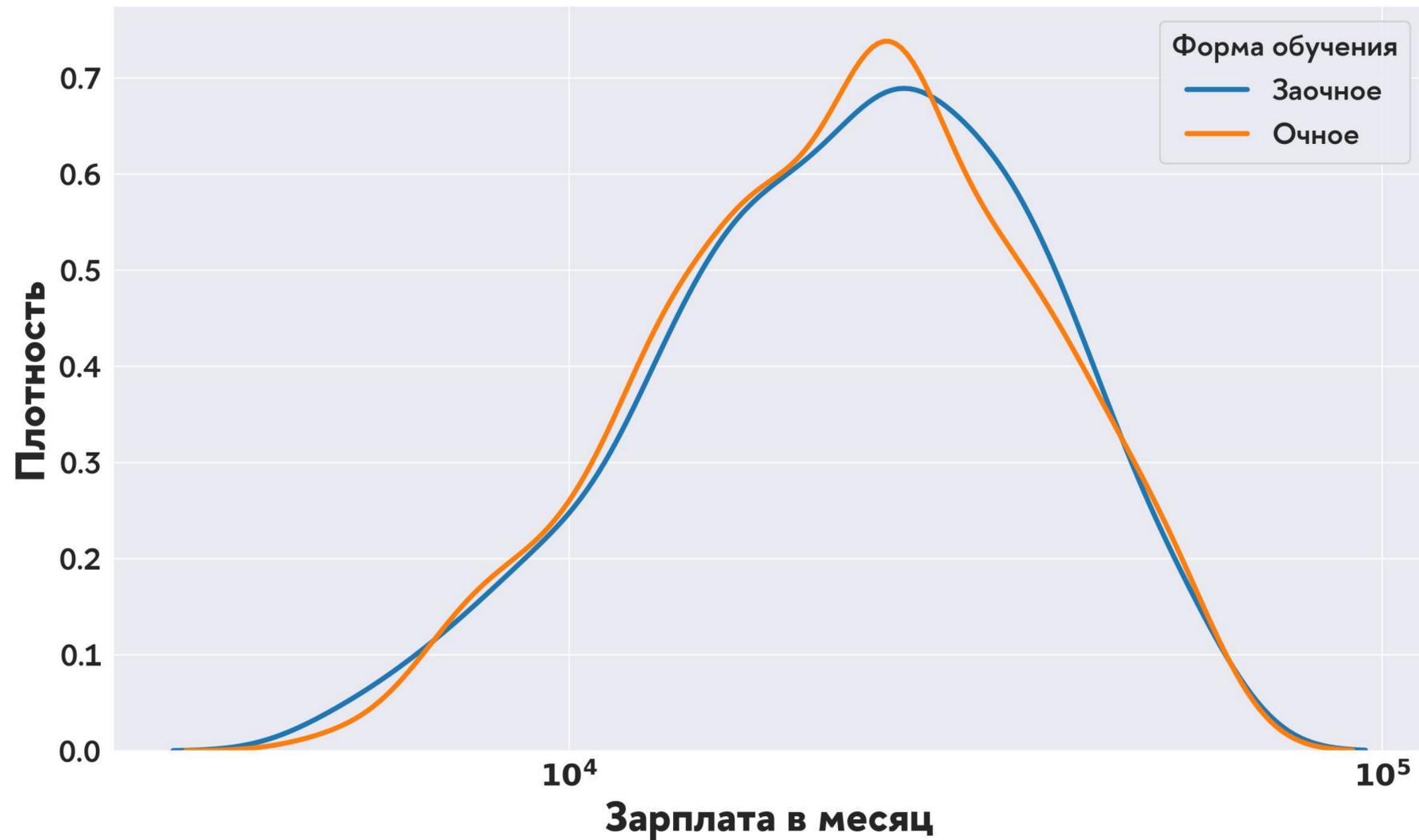
H1

Зарплаты людей, получивших образование на очной и на заочной формах образования **не равны**



U-критерий Манна-Уитни

Распределение зарплат респондентов, окончивших вуз



Результат		
Показатель	Значение	H0
p-value	0.616	✓



Проверка модели на устойчивость

Алгоритм

Разделение на группы отдельно по типам населённых пунктов и по возрастным категориям



Построение U-критерия для каждой группы



Проверка результата



Разделение на группы по населённым пунктам

Результат

Группа	Объем	p-value	H0:
Областной центр	647	0.745	✓
Город	371	0.595	✓
Село	188	0.853	✓
Посёлок городского типа	75	0.001	✗



Разделение на группы по возрастным категориям

Результат

Группа	Объем	p-value	H0:
17-25 лет	115	0.655	✓
25-45 лет	814	0.870	✓
45-60 лет	349	0.777	✓



Интерпретация

✗ Гипотеза опроверглась

Зарплаты респондентов, прошедших
очное и заочное обучение в вузе,
примерно равны

Произведена проверка устойчивости



Разделение по типам
населённых пунктов



Разделение по возрастным категориям



Ограничения и перспективы

Ограничения

Исследование не
распространяется на
ПГТ

Только 2017 год

Перспективы

Проведение более глобального опроса для исследования
влияния образования на заработную плату

Собрать большее количество данных про ПГТ, чтобы
определить, почему наши выводы здесь не работают

Провести исследование на актуальных данных, чтобы
отследить современные тенденции



Policy implication

Вузам:

Организовывать больше программ заочного обучения, так как оно менее затратно и выгодно для вузов.

Студентам:

Можно записываться на программы заочного обучения, так как они дают большее количество возможностей



Спасибо за внимание!

Остались ли вопросы?



Приложение

Распределение формата обучения

