

---

# Эффективный сотрудник - кто ты?

Команда №9

Нгуен Конг Ньат Минь      Афрамчук Владислав  
Кагиров Данис      Швец Яков      Куликова Мария

# Информация о датасете

1

Статистика по работе  
сотрудников колл-центра  
Т-Банка

- Данные по **1938** сотрудникам
- **170774** строк, **36** столбцов
- 1 сентября 2024 - 31 января 2025
- 1 запись = 1 рабочий **день**  
сотрудника
- **Дата + ID** сотрудника - уникальная  
запись

# Обработка данных

2

Удаляем технические ошибки  
**Общее** запланированное время смены  
 $\neq$   
**Сумме** плановых переменных

Удаляем записи, мешающие  
исследованию  
У сотрудника **не указан** регион

Количество записей

Было  $\rightarrow$  Стало  
**170774**  $\rightarrow$  **158817**

Количество сотрудников

Было  $\rightarrow$  Стало  
**1938**  $\rightarrow$  **1898**

# Эффективность

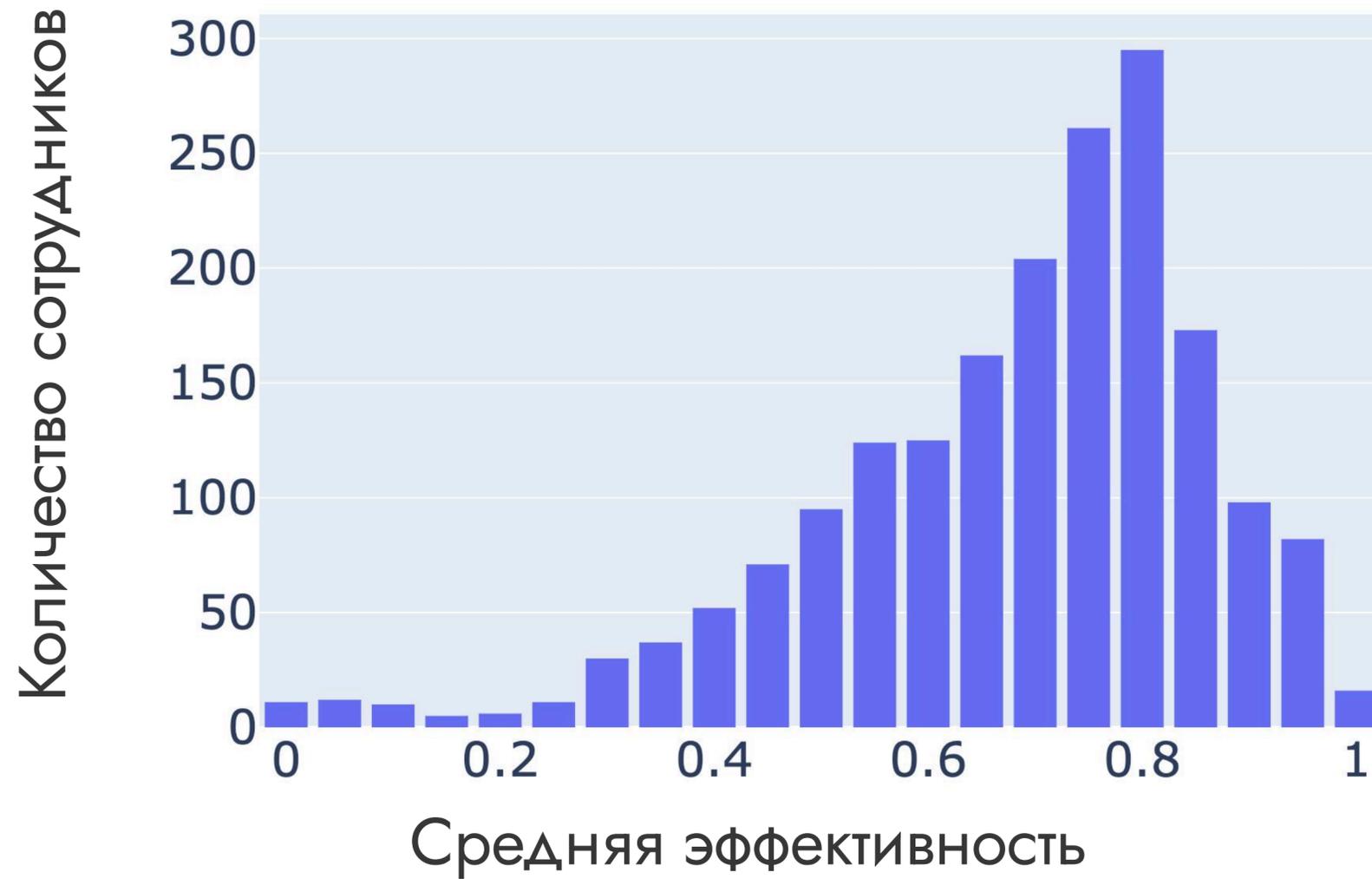
3

$$\text{Эффективность} = \frac{\text{Полезное время (звонки + чаты)}}{\text{Общее время в системе}}$$

Эффективность показывает долю времени, в течение которого сотрудник непосредственно приносил результат.

# Средняя эффективность

4



Среднее:

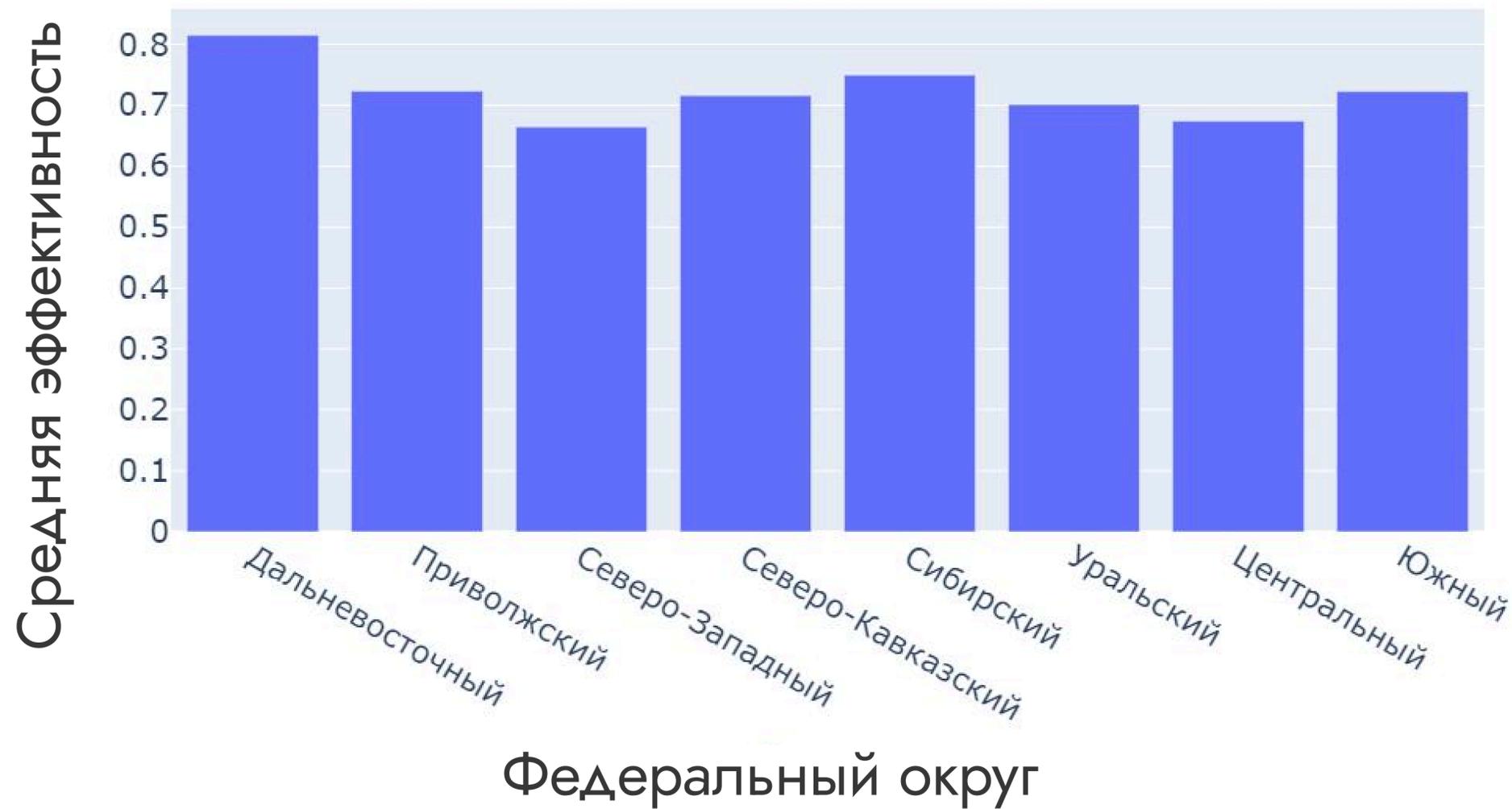
**0.681**

Медиана:

**0.722**

# Регионы

5

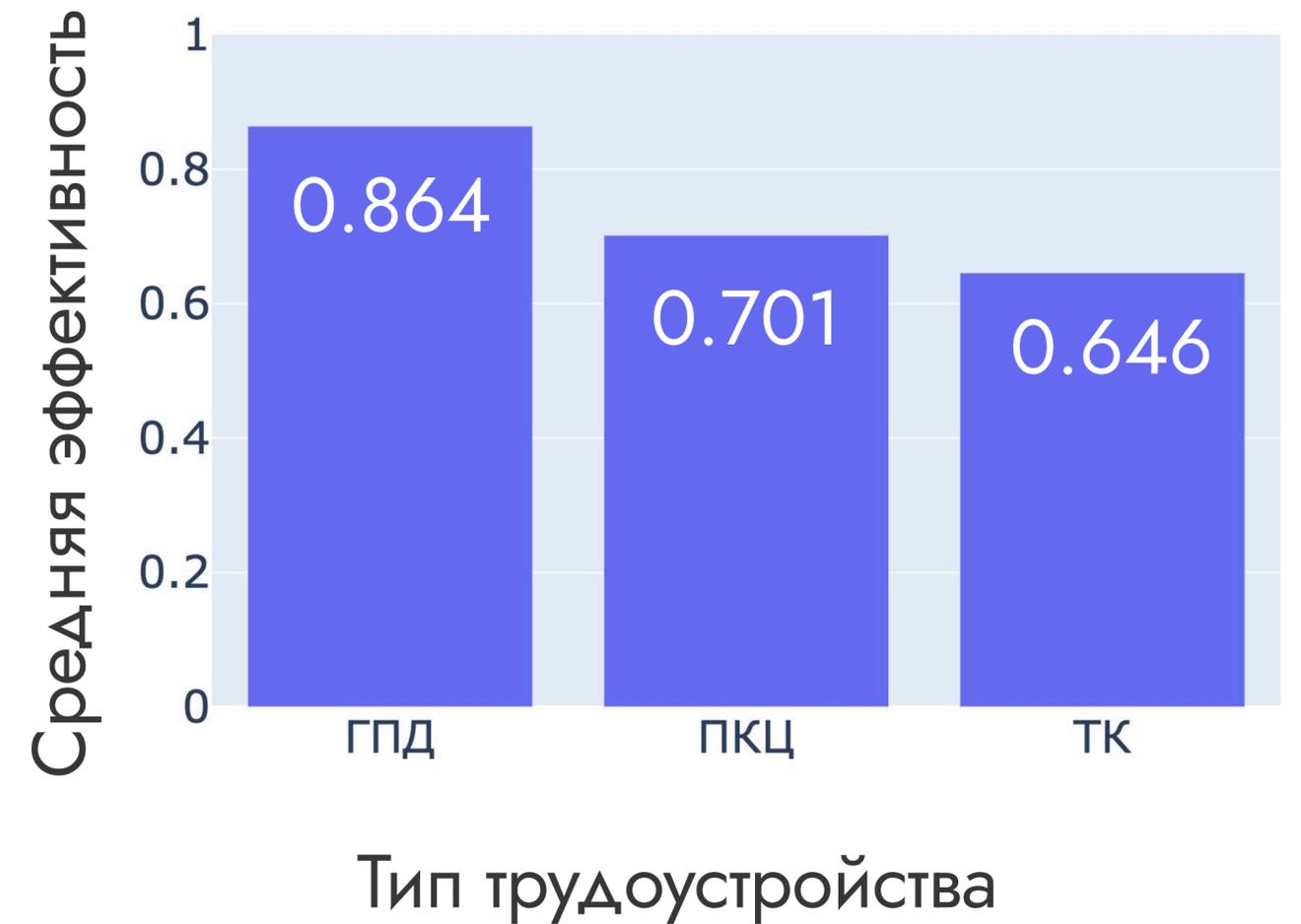


Новая переменная

Округ

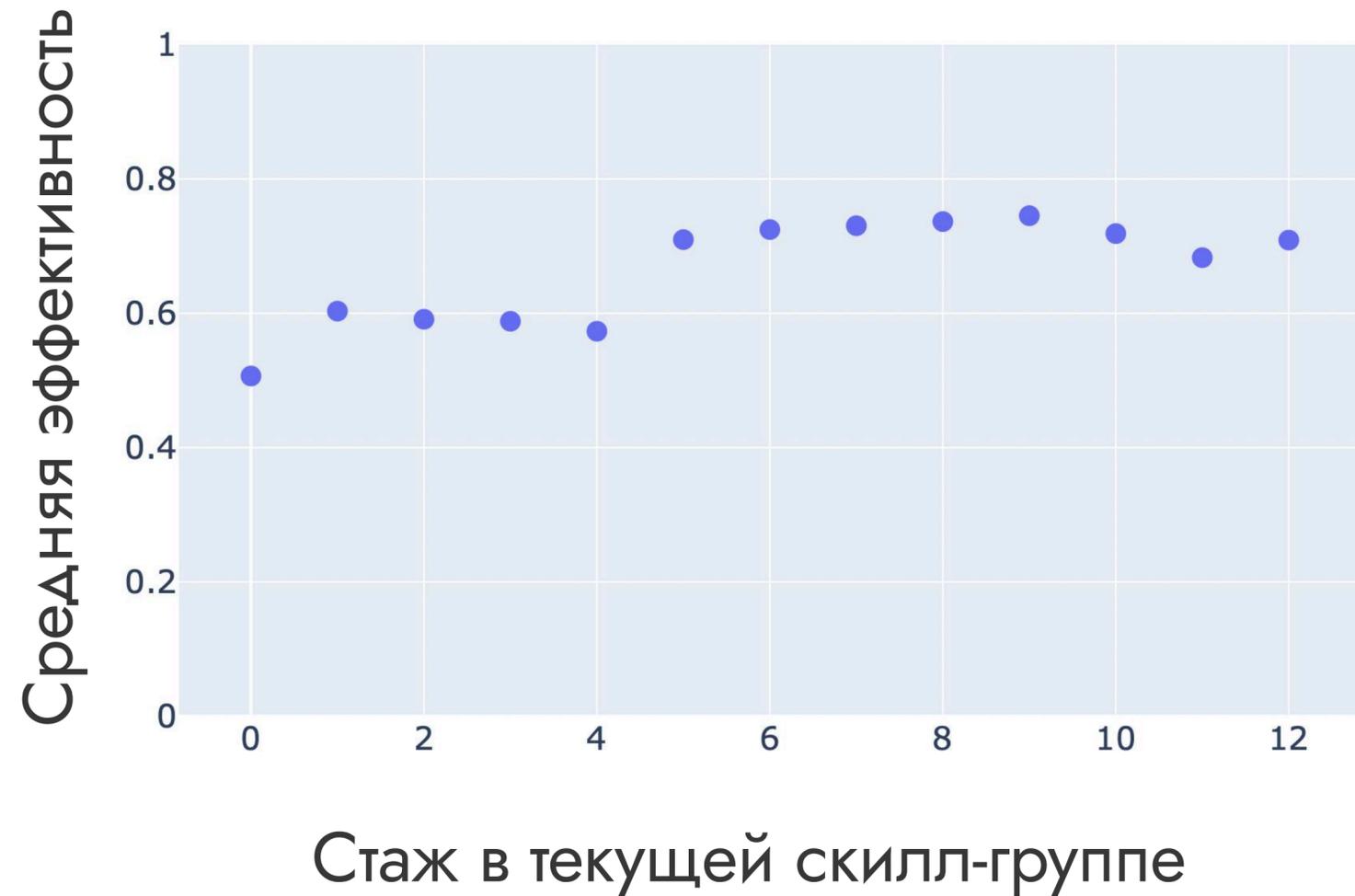
# Тип трудоустройства

6



# Совместное распределение

7



Коэффициент  
корреляции

**0.903**

# Исследовательский вопрос и гипотеза

8

## Исследовательский вопрос



Как выглядит портрет эффективного сотрудника?

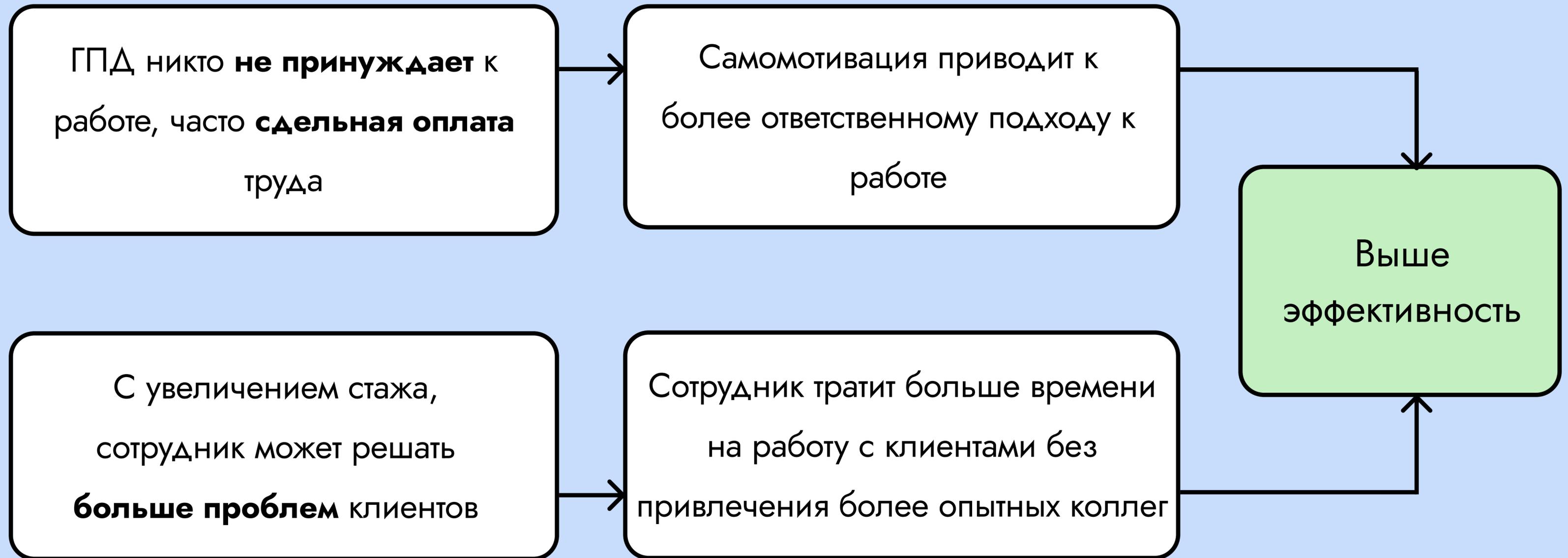
## Гипотеза



Сотрудники с типом трудоустройства **ГПД** и наибольшим **стажем** в скилл-группе показывают наибольшую эффективность

# Механизм

9



# Математическая модель

10

Критический уровень статзначимости: **0.05**

Гипотеза о ГПД

Сравнение средних эффективностей  
сотрудников ГПД с остальными подгруппами  
с помощью **Z-теста**

Гипотеза о стаже

Регрессионный анализ  
Зависимая переменная: **Эффективность**  
Независимая переменная: **Стаж** в скилл группе

# Проверка гипотезы о ГПД

11

**H0:** Средняя эффективность сотрудников обеих групп одинакова

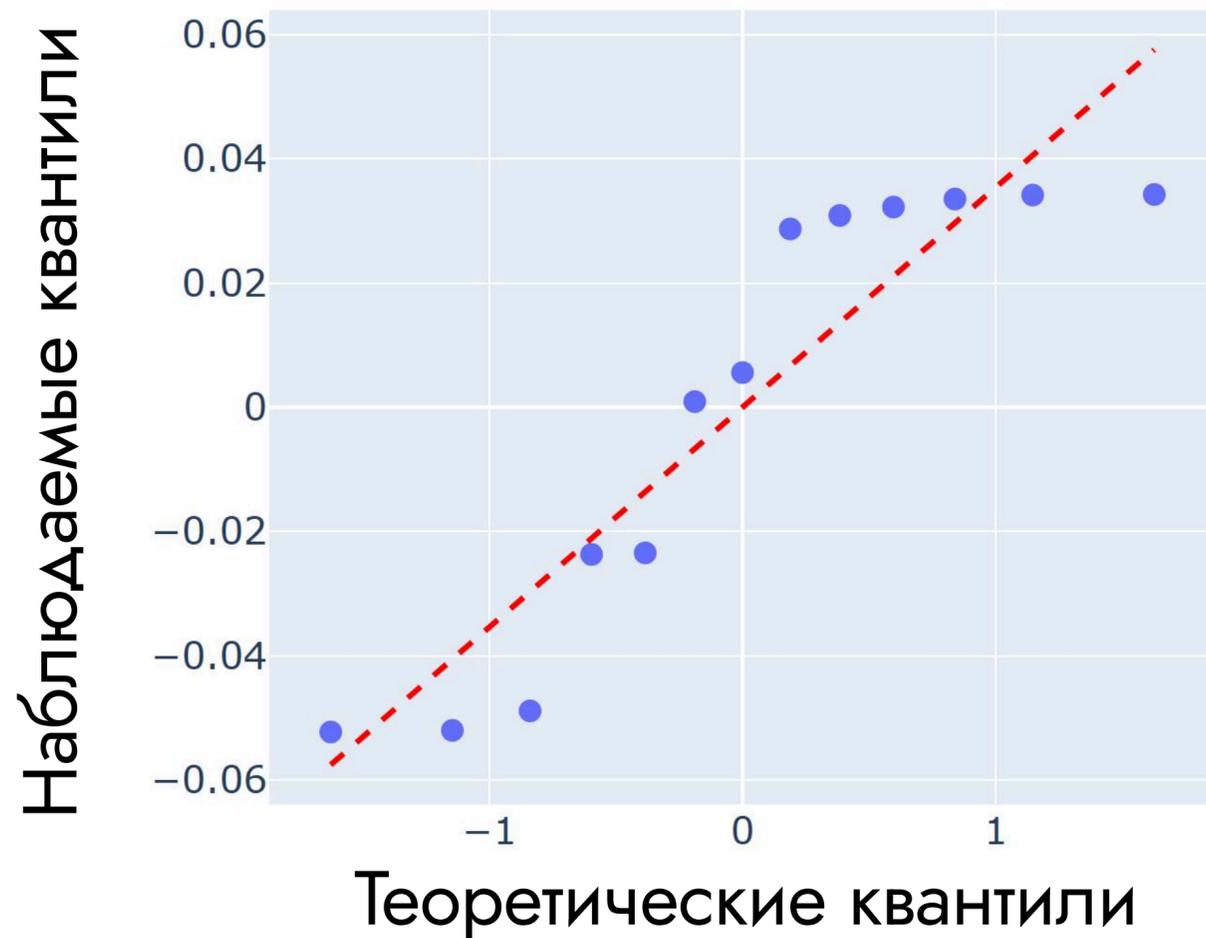
**H1:** Средняя эффективность сотрудников различается 

Группы сравнения	Размеры выборок	Z-статистика	p-value
ГПД и ТК	262 и 1583	25.0092	0.000
ГПД и ПКЦ	262 и 53	7.2993	0.000

# Требования линейной регрессии

12

## Q-Q plot



## Тест Шапиро-Уилка

Статистическая оценка

**0.82**

p-value

**0.012**

# Требования линейной регрессии

13

## Проблема

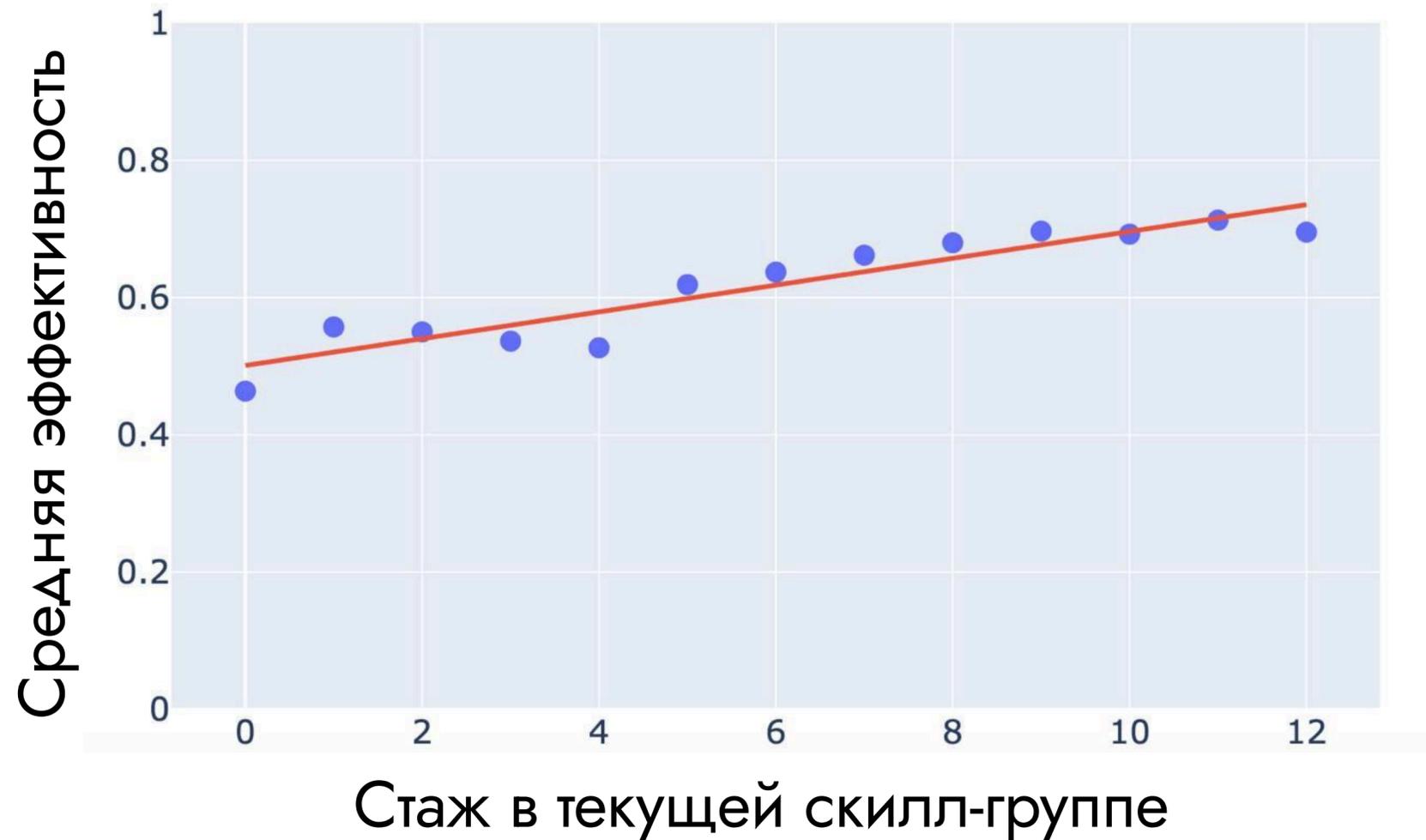
Распределение остатков нельзя назвать нормальным, линейную регрессию использовать нельзя.

## Решение

Вместо нее воспользуемся робастной регрессией, которая не требует многих предпосылок линейной регрессии.

# Робастная регрессия

14



$$r = \mathbf{0.936}$$

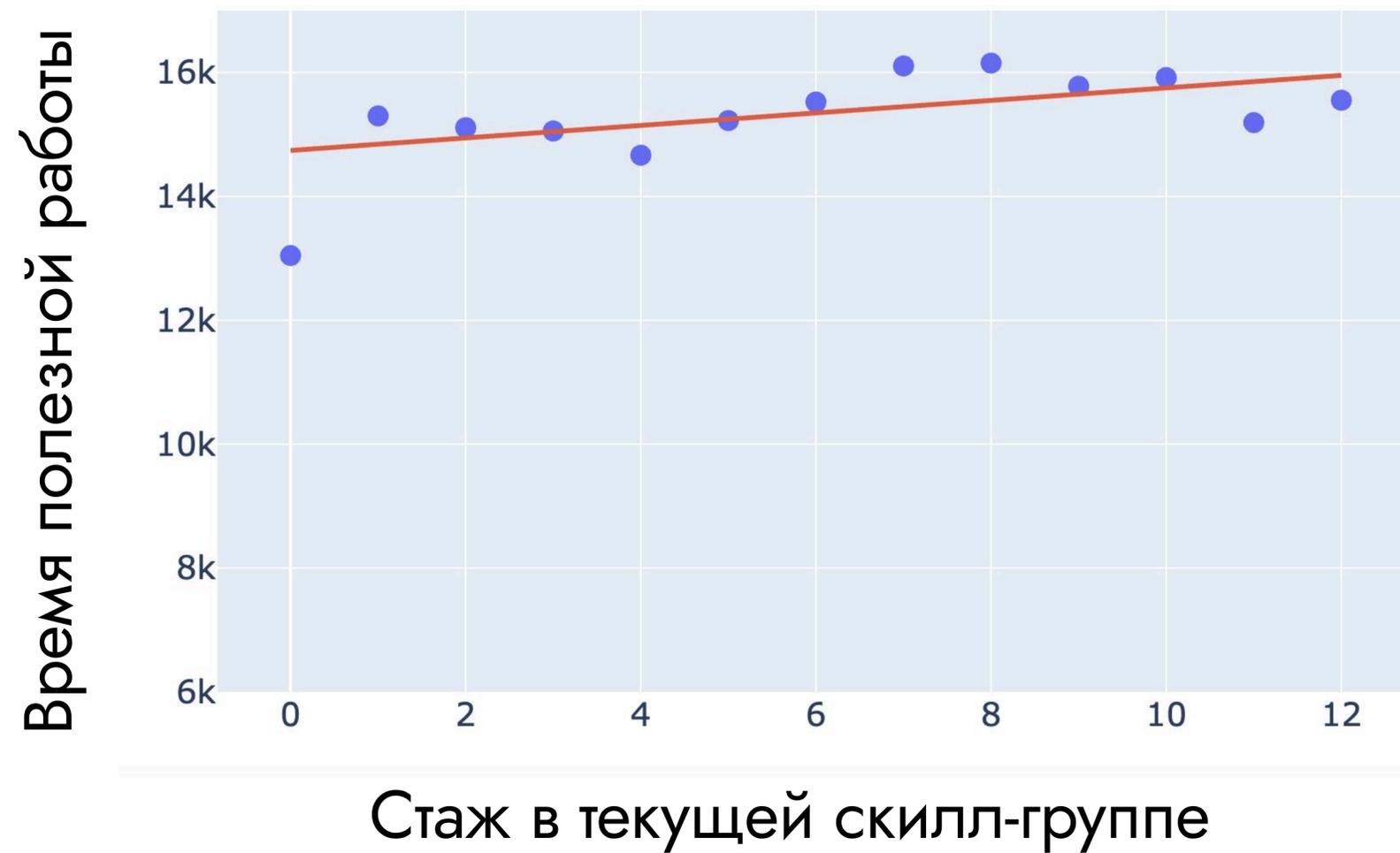
$$R^2 = \mathbf{0.875}$$

$$Z\text{-score} = \mathbf{8.249}$$

$$p\text{-value} = \mathbf{0.000}$$

# Подтверждение механизма

15



$r = 0.64$

$R^2 = 0.38$

Z-score = **2.57**

p-value = **0.01**

# Проверка устойчивости

16

Проверка по группам

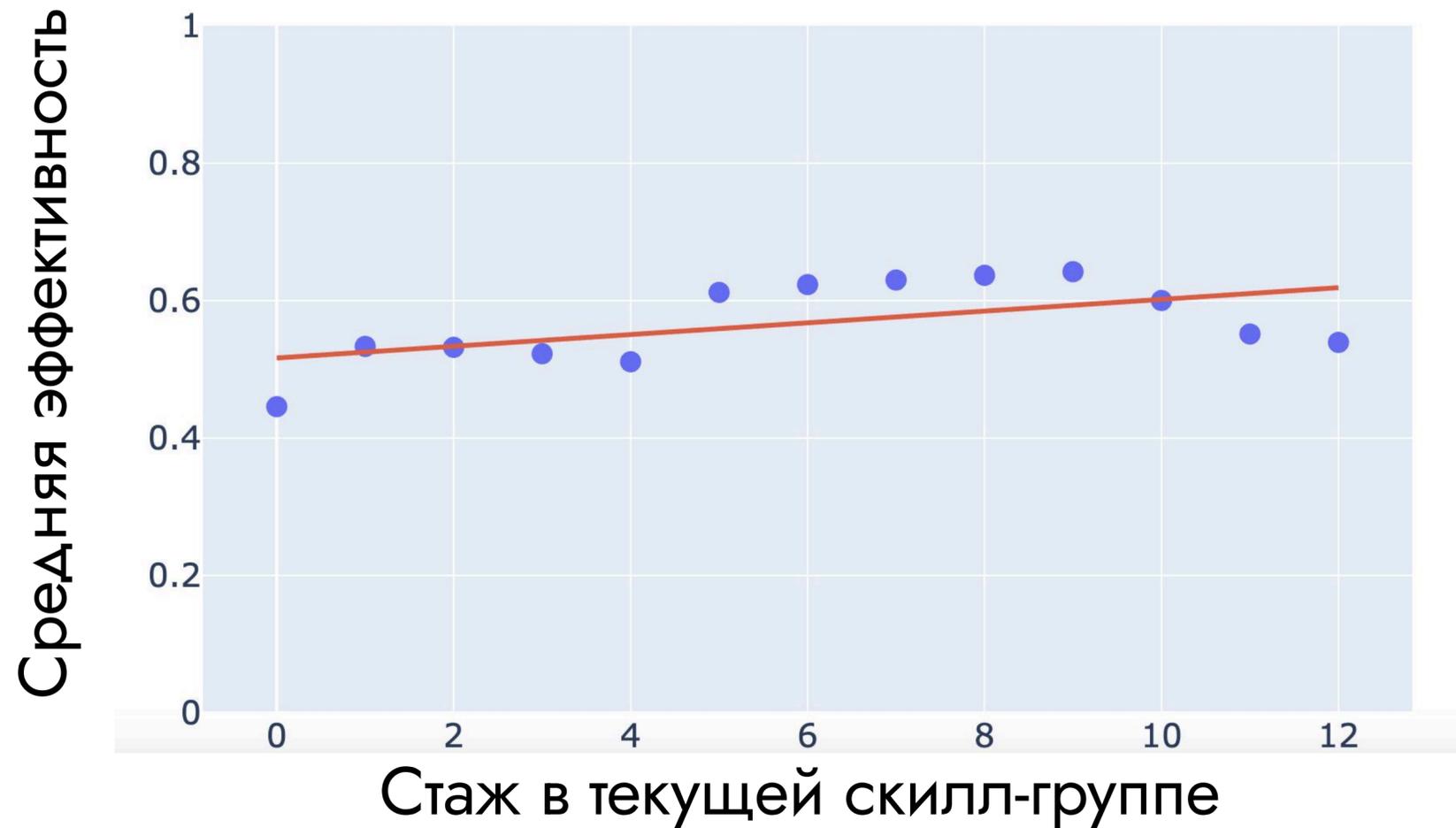
Разбиваем по **типам трудоустройства** и **федеральным округам**

**В каждой группе:**

1. Строим робастную регрессию
2. Анализируем результат

# Устойчивость по типам трудоустройства (ТК)

17



Размер выборки:

**143 330**

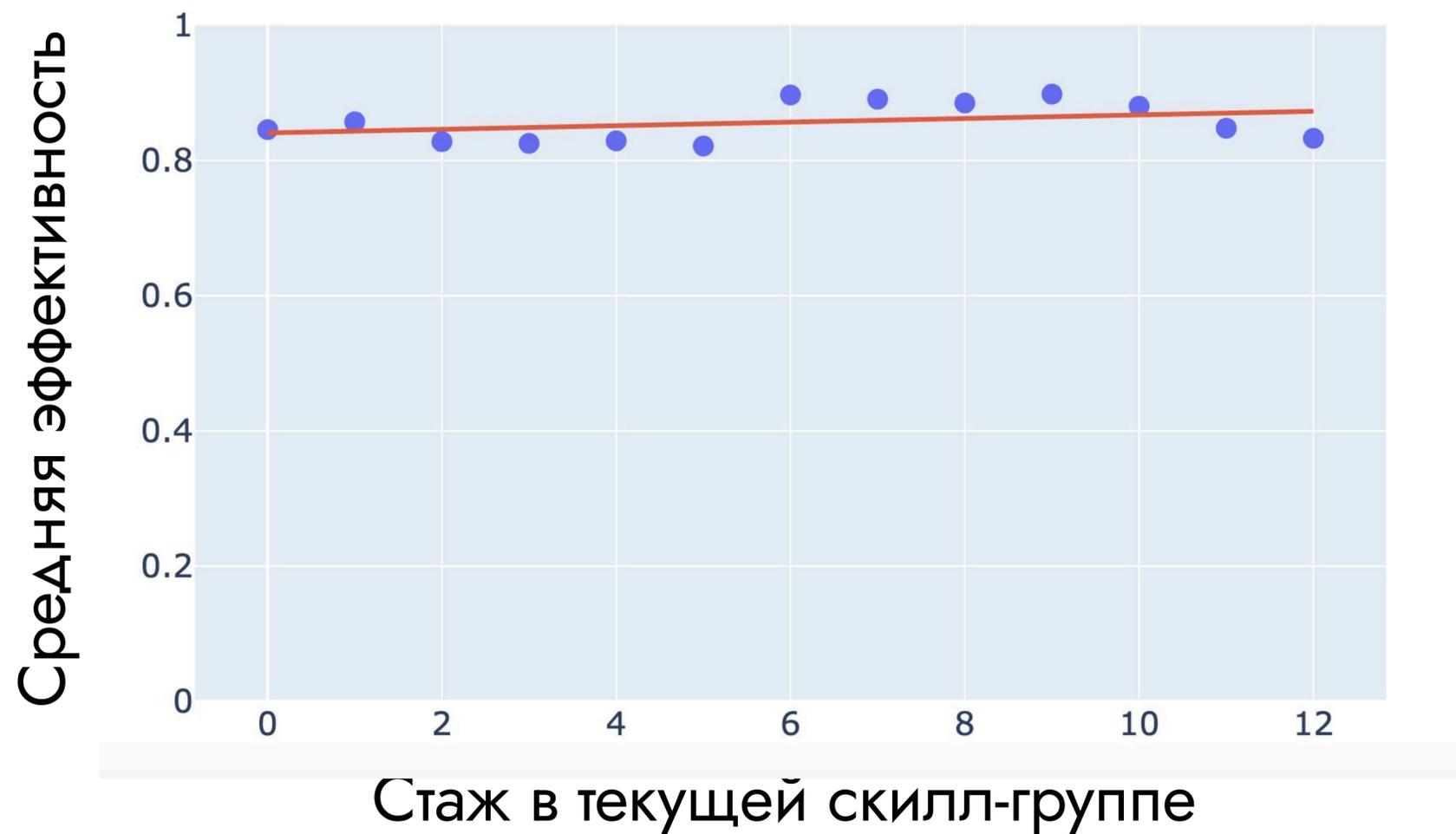
$r = 0.549$

$R^2 = 0.301$

Z-score = **2.177**

p-value = **0.03**

# Устойчивость по типам трудоустройства (ГПД)



Размер выборки:

**22 686**

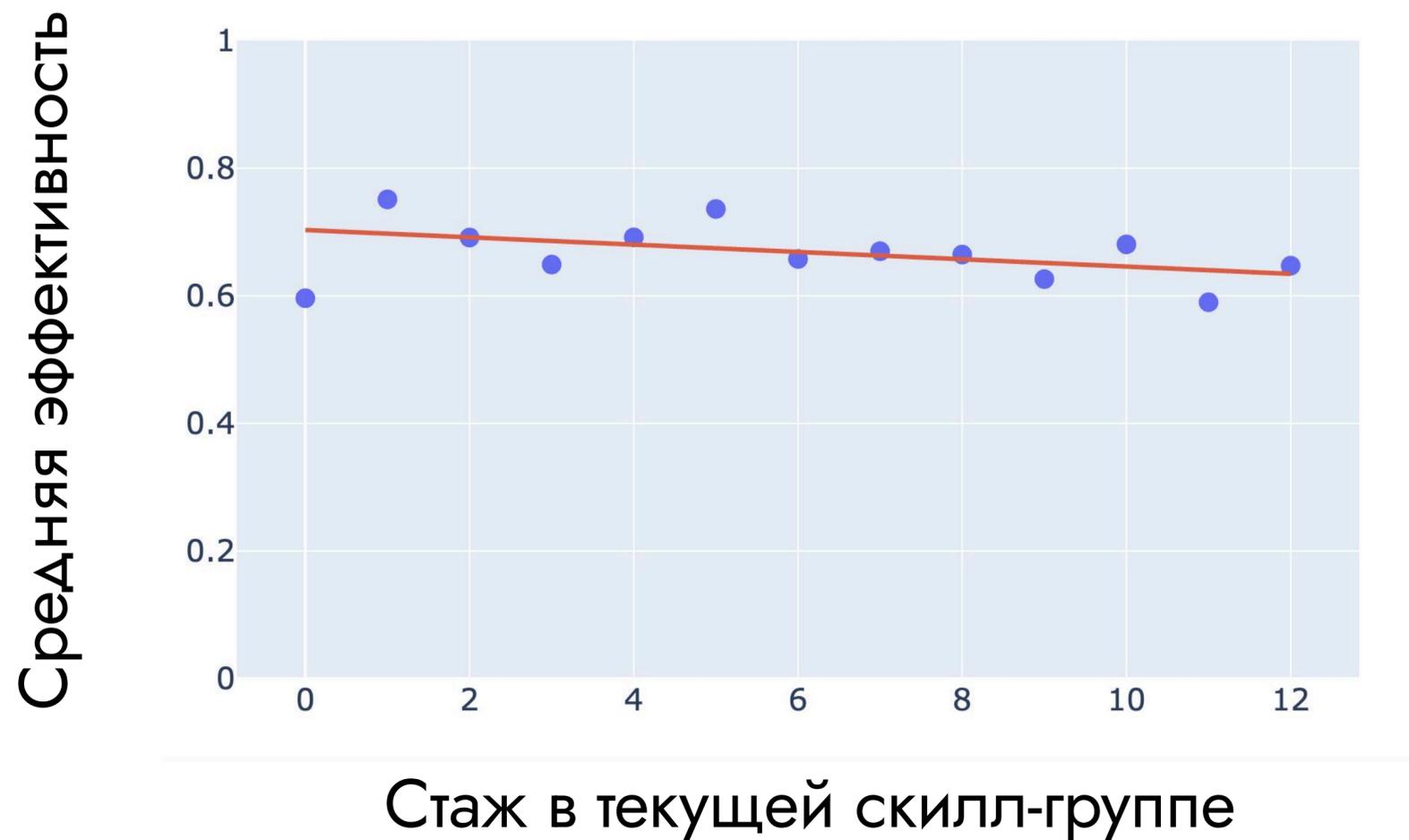
$r = 0.351$

$R^2 = 0.123$

Z-score = **1.246**

p-value = **0.213**

# Устойчивость по типам трудоустройства (ПКЦ)



Размер выборки:  
**2222**

$r = -0.332$

$R^2 = 0.08$

Z-score = **-1.597**

p-value = **0.11**

# Устойчивость по округам, Робастная регрессия

20

Федеральный округ	Размер выборки	R <sup>2</sup>	p-value
ЮФО	20212	0.94	0.000
ЦФО	64189	0.71	0.000
СЗФО	20261	0.9	0.000
СКФО	2744	0.03	0.589
ПФО	27261	0.83	0.000
УФО	20212	0.75	0.000
СФО	11050	0.72	0.000
ДФО	667	0.04	0.363

# Интерпретация результатов

21



Гипотеза подтвердилась частично

В среднем сотрудники с типом трудоустройства ГПД более эффективны, однако получение опыта влияет на них не так сильно.



Проверка на устойчивость

Модель устойчива на большинстве федеральных округов  
Влияние стажа зафиксировано лишь для ТК операторов

# Альтернативный механизм

22

С увеличением стажа,  
сотрудник **быстрее**  
**выполняет**  
запланированную  
работу

Общее время  
сотрудника в  
системе  
**уменьшается**

Выше  
эффективность

# Практическая польза

23

В случае найма новых сотрудников сфокусироваться на ГПД

При работе с ТК не переводить их часто на другую скилл-группу, давать им набираться опыта в текущей скилл-группе

# Ограничения

24

Метрика эффективности не в полной мере оценивает работу оператора

Недостаточное кол-во операторов для проверки на подвыборках

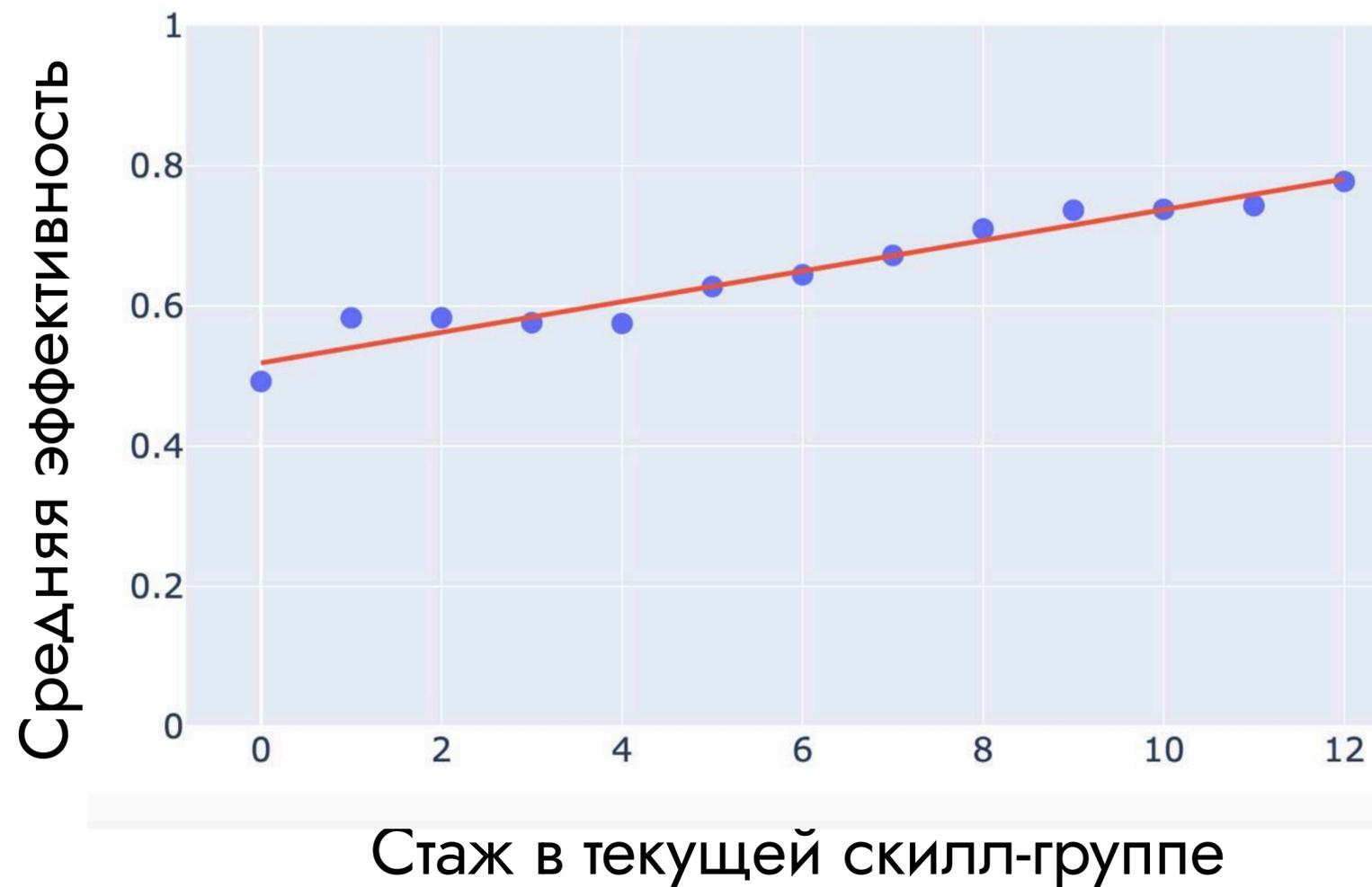
Нет данных о стаже работы операторов до устройства в компанию

**Спасибо за внимание!**

Можете задать интересующие вас вопросы

# Робастная регрессия (ЮФО)

25



$r = 0.97$

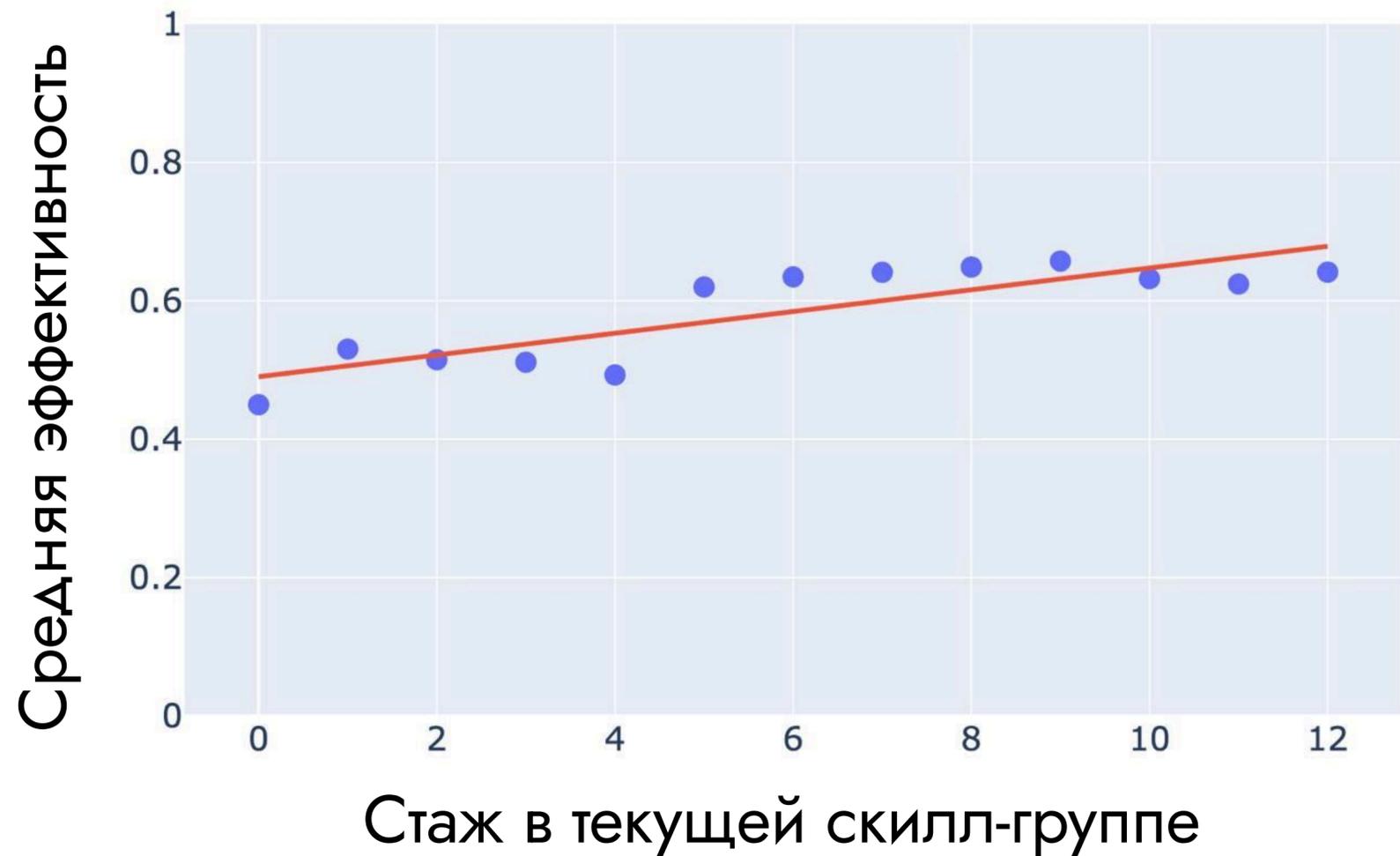
$R^2 = 0.95$

Z-score = 13.57

p-value = **0.000**

# Робастная регрессия (ЦФО)

26



$r = 0.84$

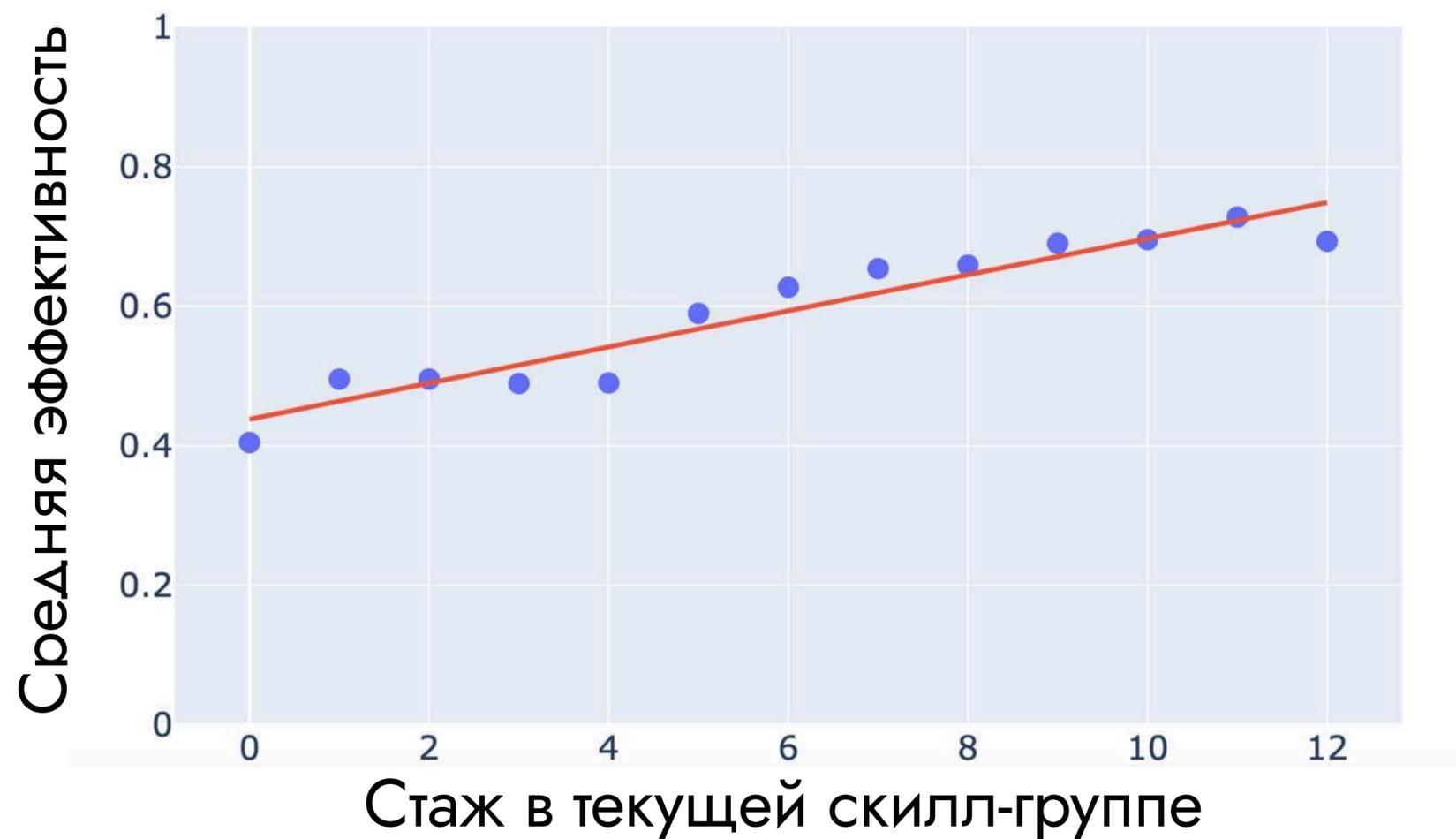
$R^2 = 0.73$

Z-score = 5.2

p-value = **0.000**

# Робастная регрессия (СЗФО)

27



$r = 0.95$

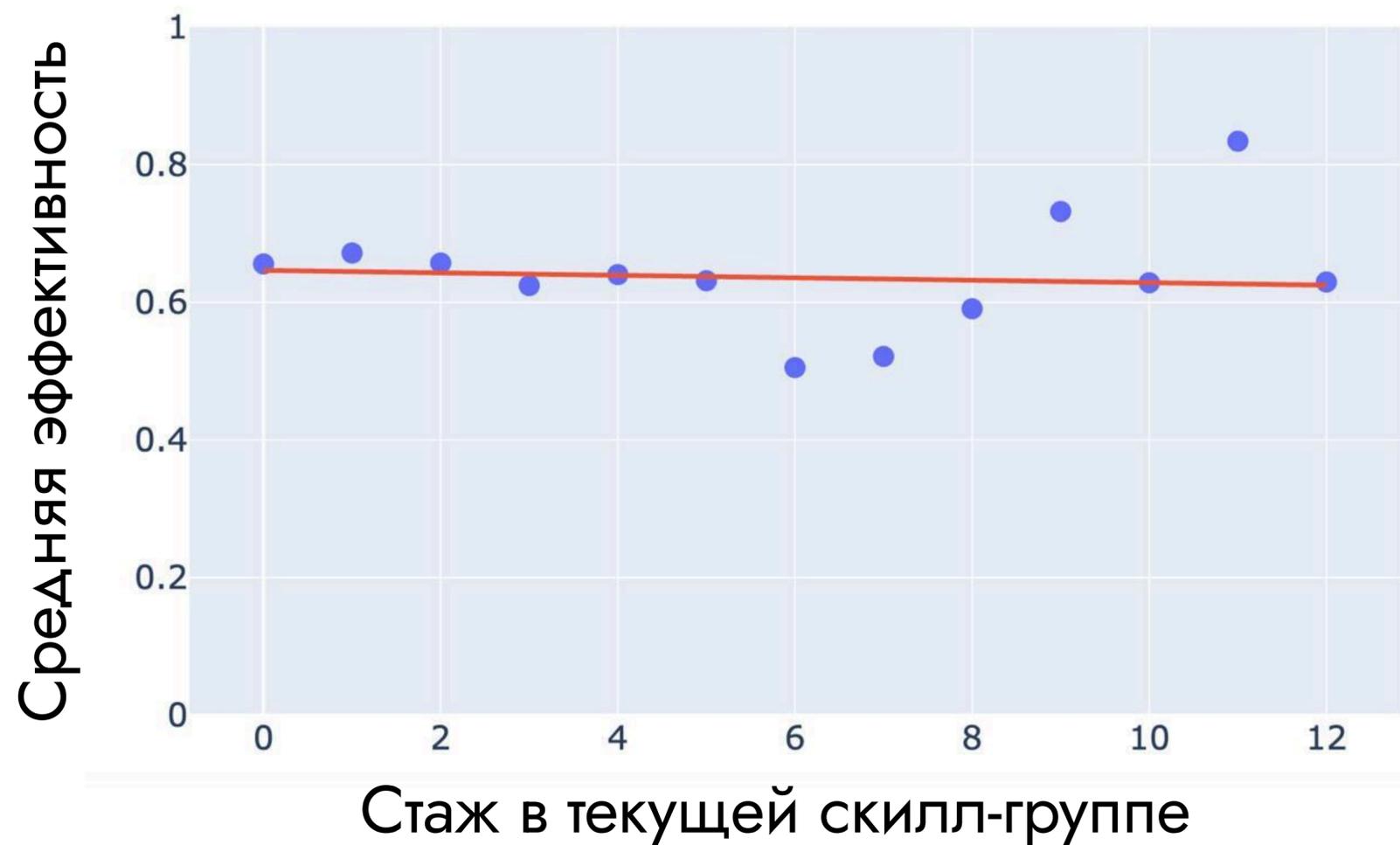
$R^2 = 0.91$

Z-score = 9.75

p-value = **0.000**

# Робастная регрессия (СКФО)

28



$r = -0.16$

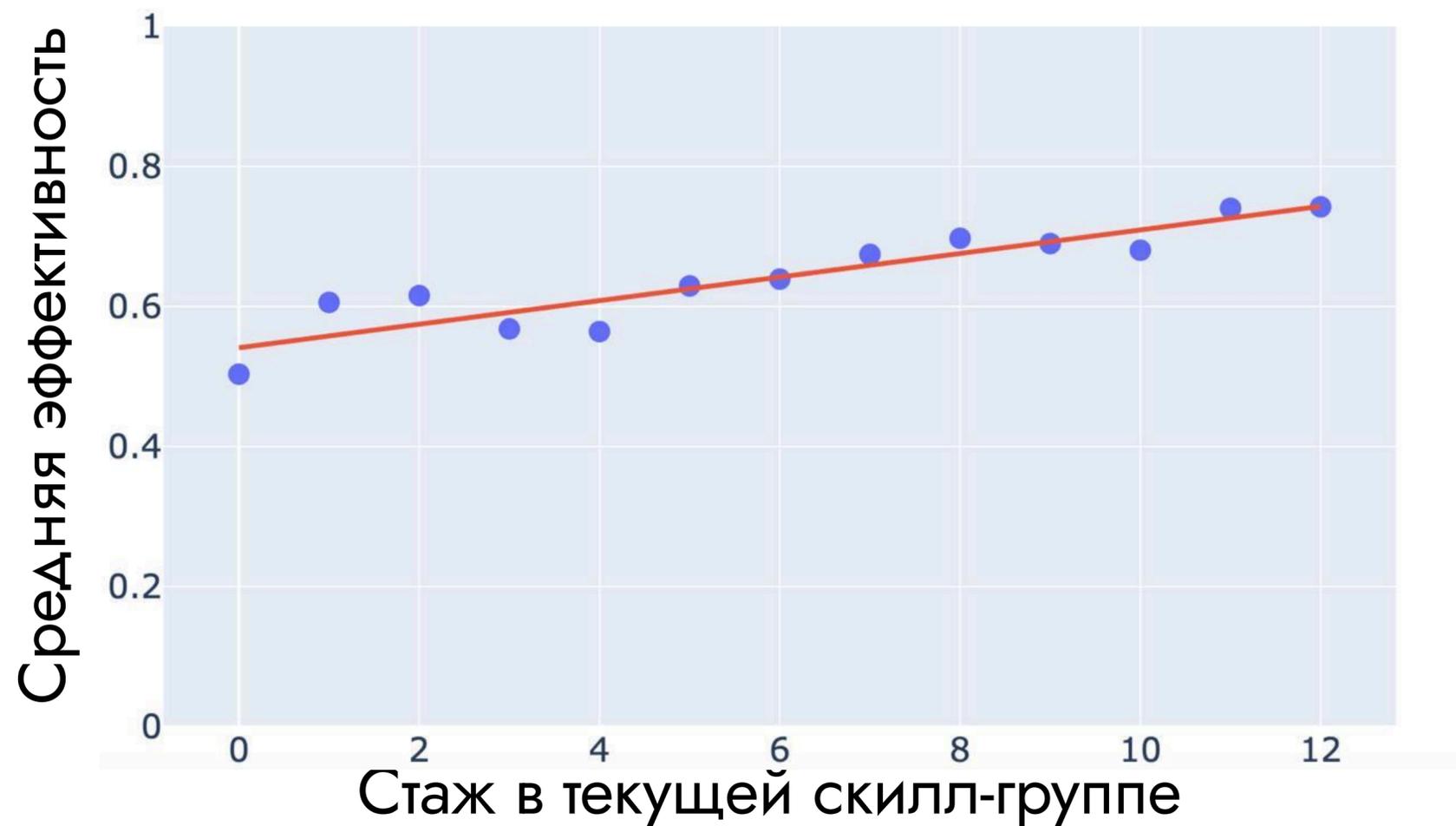
$R^2 = 0.03$

Z-score = -0.54

p-value = **0.589**

# Робастная регрессия (ПФО)

29



$r = 0.92$

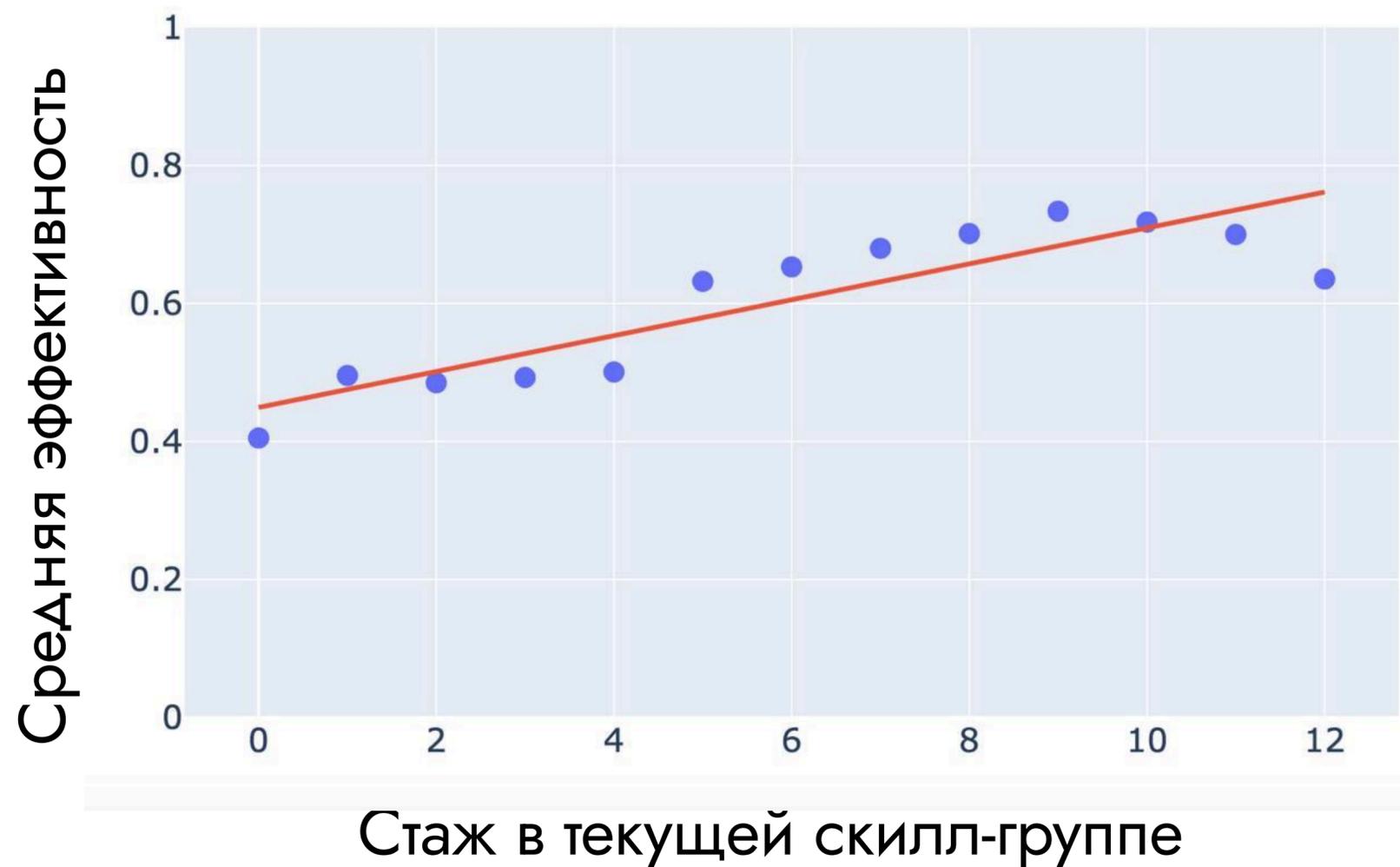
$R^2 = 0.84$

Z-score = 6.47

p-value = **0.000**

# Робастная регрессия (УФО)

30



$r = 0.87$

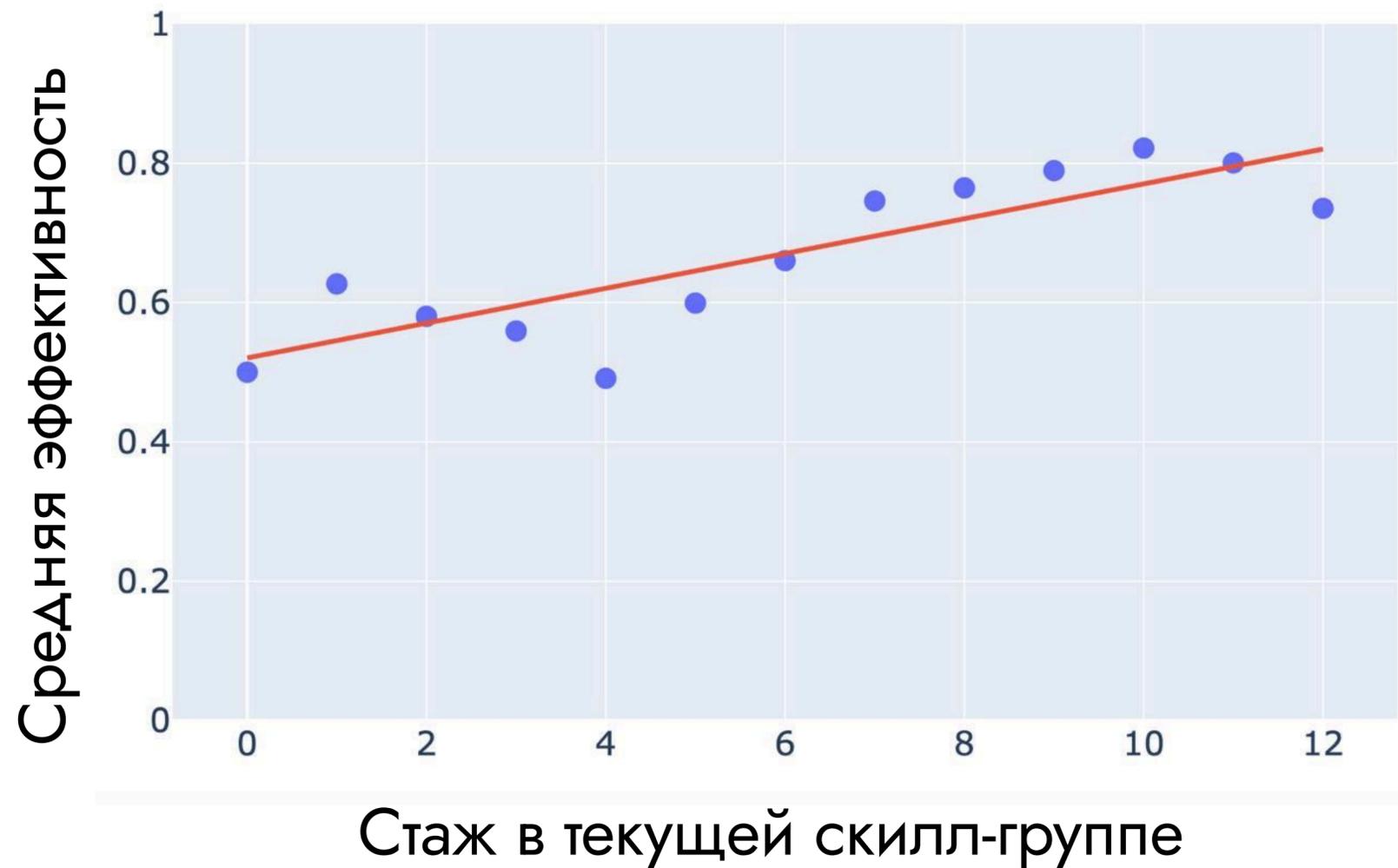
$R^2 = 0.76$

Z-score = 6.38

p-value = **0.000**

# Робастная регрессия (СФО)

31



$r = 0.85$

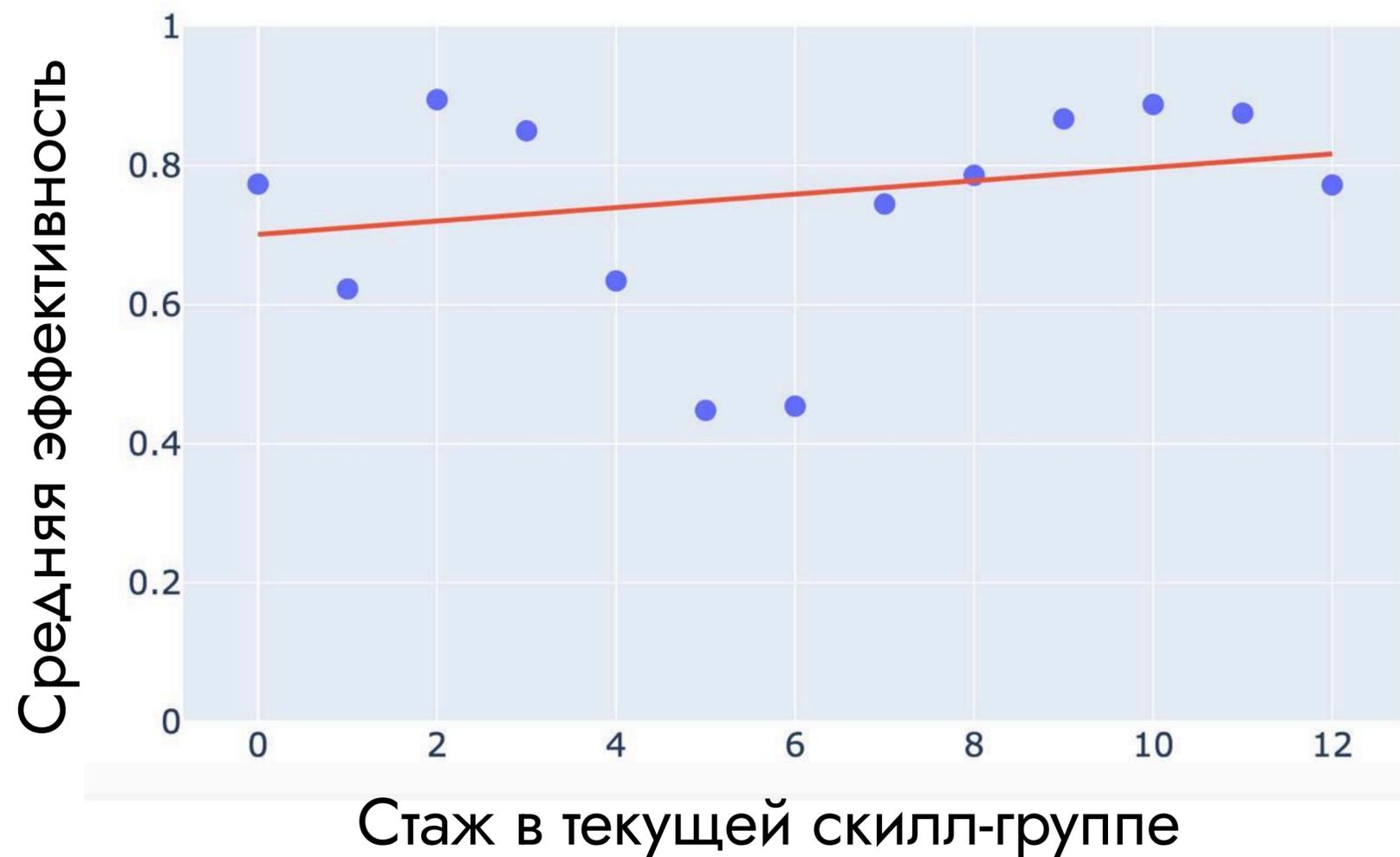
$R^2 = 0.73$

Z-score = 5.46

p-value = **0.000**

# Робастная регрессия (ДФО)

32



$r = 0.26$

$R^2 = 0.05$

Z-score = 0.90

p-value = **0.363**

# Q-Q Plot на всем датасете

