

ШТРАФЫ И ДЕНЬГИ

14

Братухин Арсений

Степанов Роберт

Возяков Вячеслав

Шаматова Дарья

Димитриева Мария

Коваленко Андрей

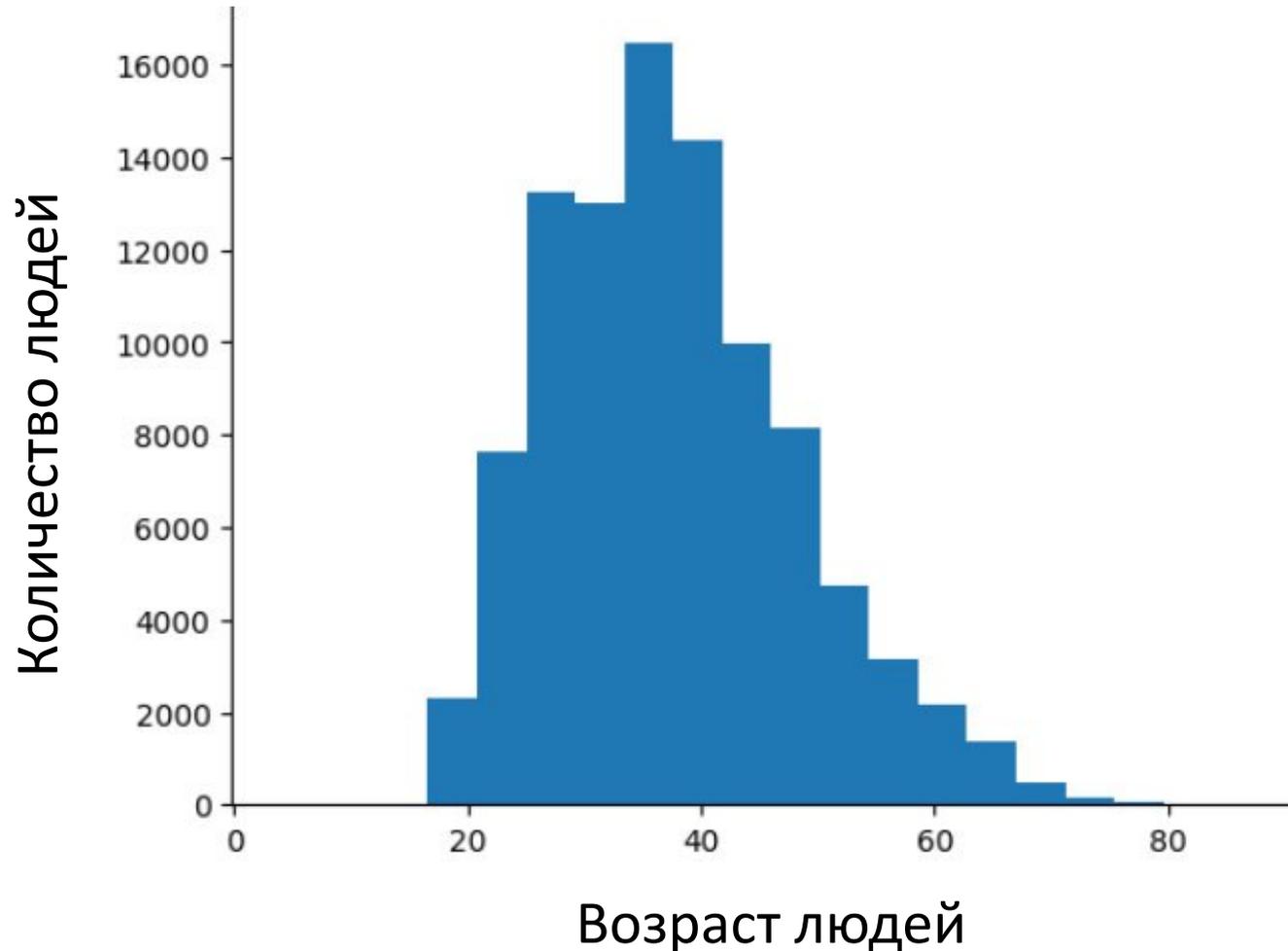
Датасет и его структура

- 🚗 Штрафы ГИБДД: инструменты регулирования поведения участников дорожного движения
- 🚗 22 показателя
- 🚗 Данные за 1 месяц
- 🚗 84 субъекта РФ
- 🚗 Распределение большинства параметров ненормальное

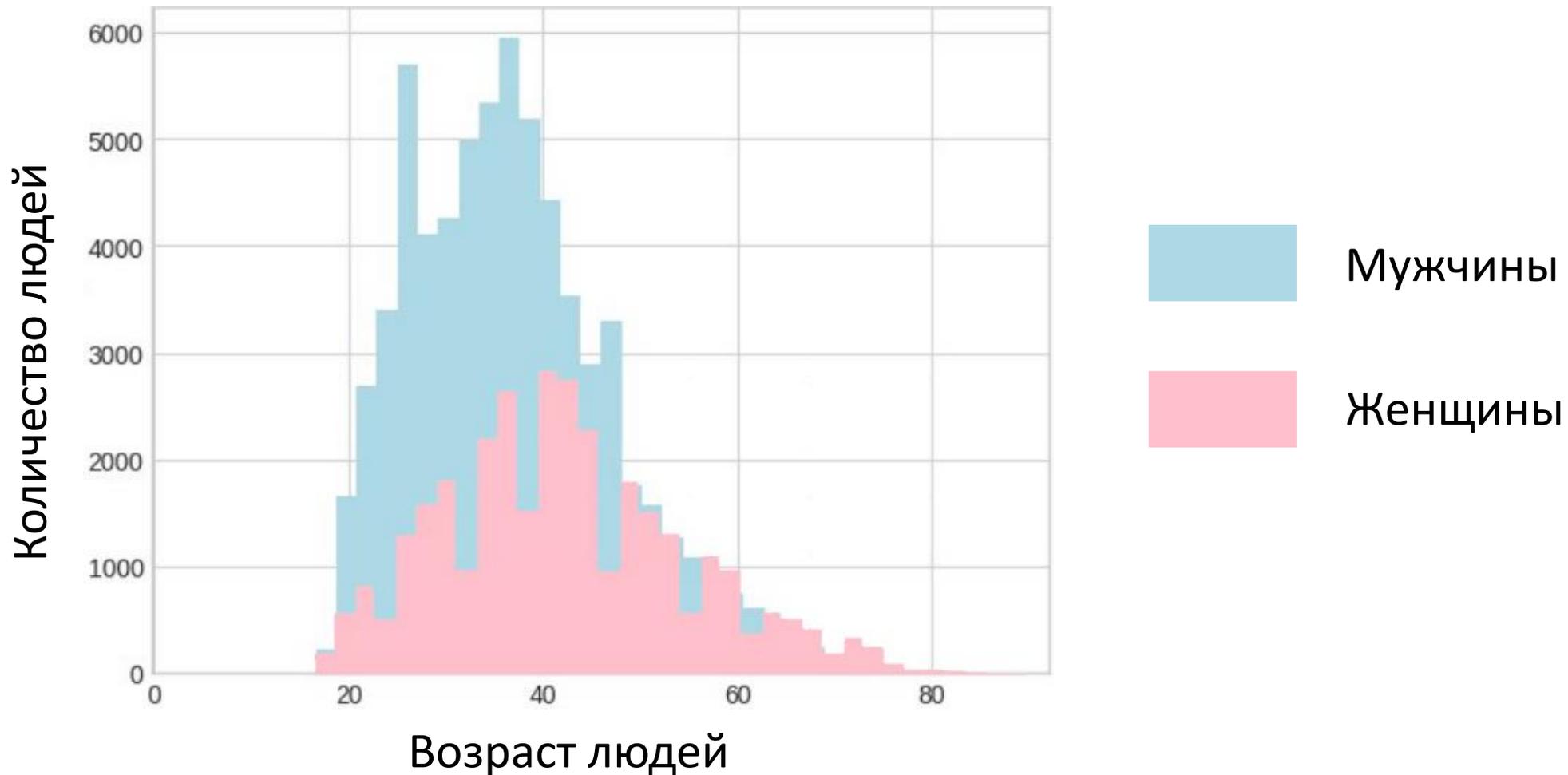


Распределение основных величин

Распределение людей по возрасту

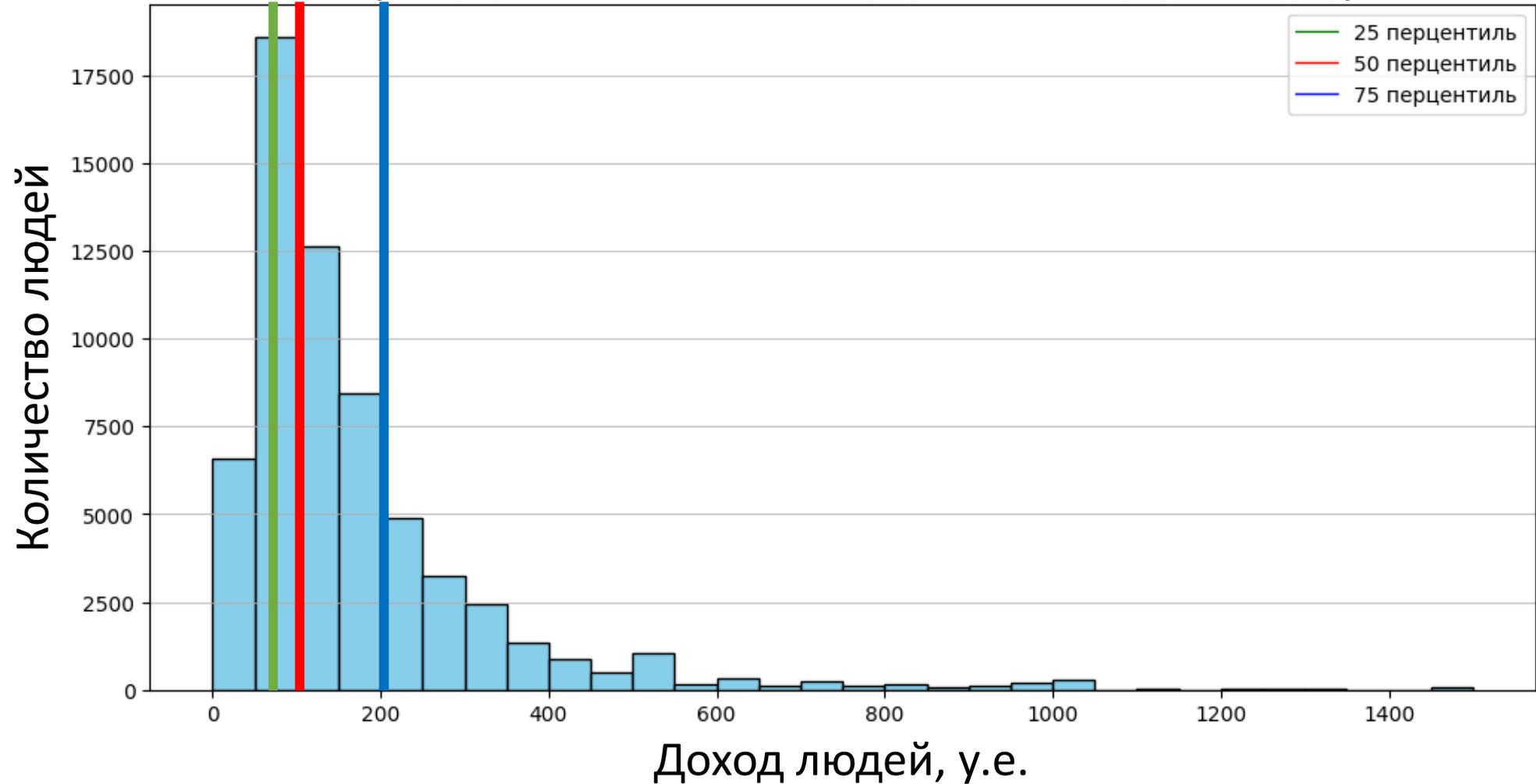


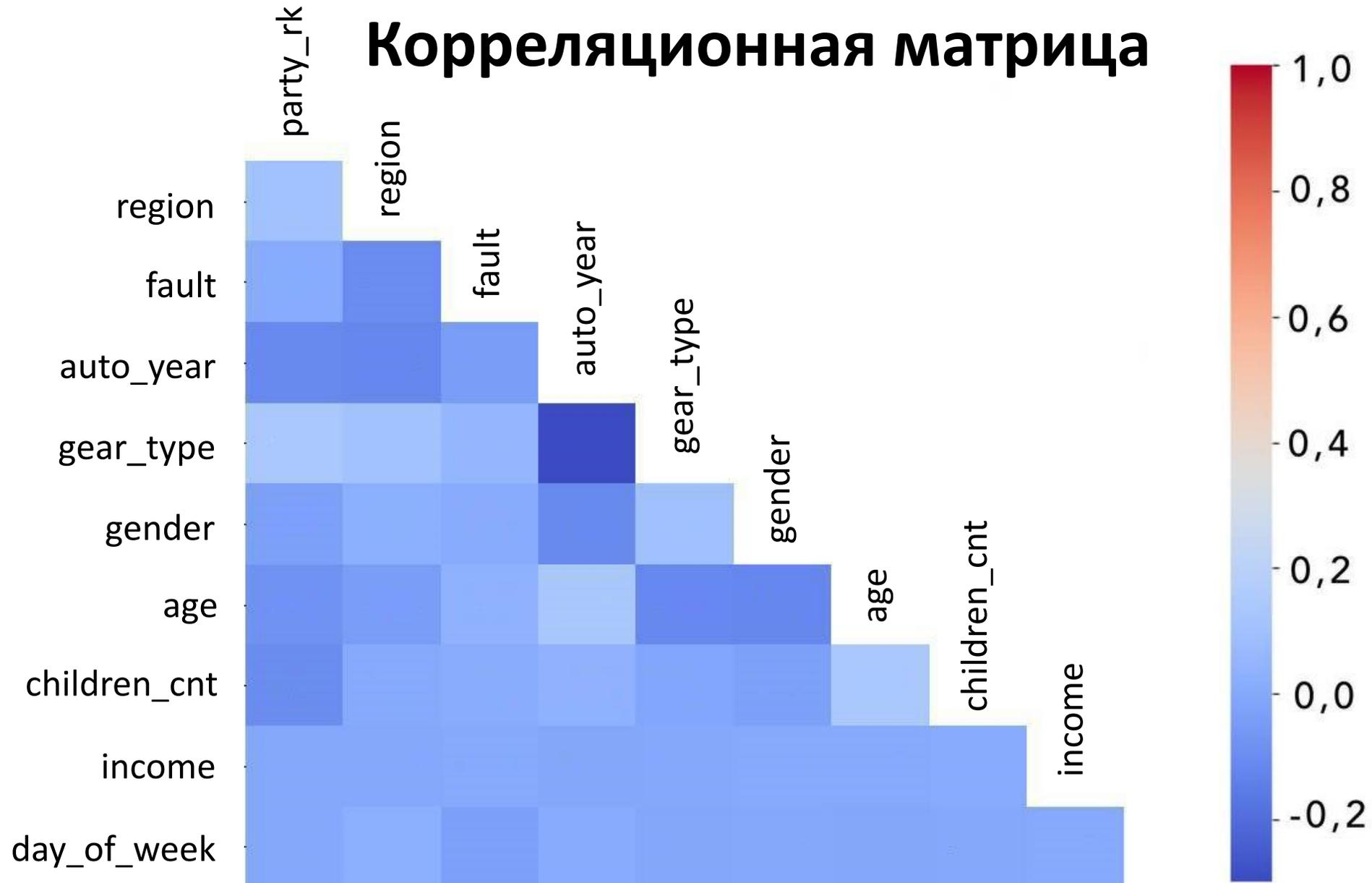
Распределение основных величин



Распределение основных величин

Распределение людей по доходу





Исследовательский вопрос

Взаимосвязаны ли характеристики клиентов и их автомобилей с характеристиками правонарушений?

Н0

Нарушители с высоким доходом имеют меньшее количество штрафов по сравнению с нарушителями из других доходных групп.

Н1

Нарушители с высоким доходом имеют не меньшее количество штрафов по сравнению с нарушителями из других доходных групп.



Теория

Предвар. анализ

Гипотеза

Проверка гипотезы

Выводы

Механизмы

Механизм Н0

У богатых людей дорогие машины

Их ремонт стоит дороже

Они боятся их разбить

Реже нарушают ПДД

Получают меньше штрафов

Механизм Н1

У богатых людей больше свободных денег*

Оплата штрафов для них не обременительна

Они могут позволить себе мелкие правонарушения

Чаще нарушают ПДД

Получают больше штрафов

**Будем называть деньги свободными, если они не задействуются для оплаты базовых потребностей человека*

Подготовка данных

- ☞ Проверили наличие выбросов и удалили их
- ☞ Удалили явно некорректные данные
- ☞ Посчитали количество штрафов для каждого человека по его party_rk

engine_type

Бензиновый на бензине

engine_power

окт.00

car_price

;M

age

4

Мат. модель

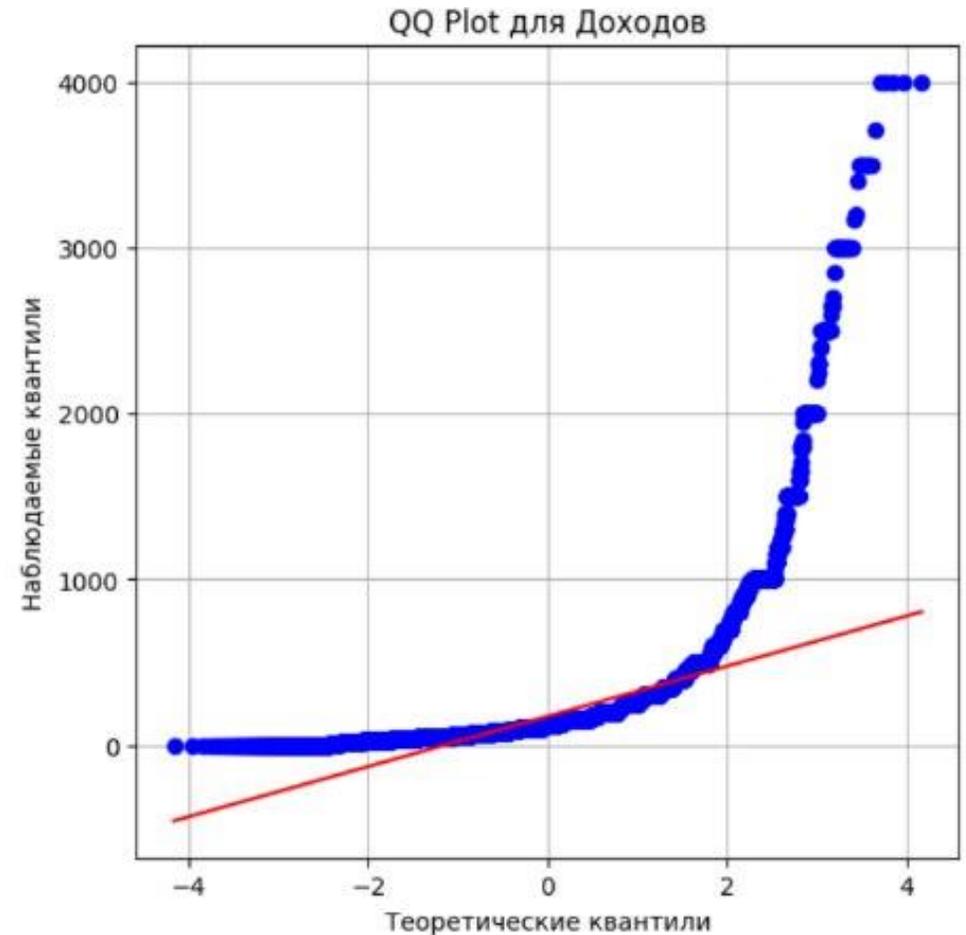
H0: Распределение доходов похоже на нормальное

H1: Распределение доходов не является нормальным

Критерий Шапиро-Уилка:

р-значение: $1.86e-154$

**Распределение доходов не является нормальным
(отвергаем H0)**



Мат. модель

Критерий Краскела-Уоллиса для всего датасета:

p-значение: $2.12e-42$

H0: нет статистически значимых различий в штрафах между группами по доходам

H1: есть статистически значимые различия в штрафах между группами по доходам

Есть статистически значимые различия в штрафах между группами по доходам (отвергаем H0)



Доход связан с количеством штрафов

Мат. модель

по полу

Критерий Краскела-Уоллиса для женщин:
p-значение: $1.45e-15$

Критерий Краскела-Уоллиса для мужчин:
p-значение: $7.87e-25$

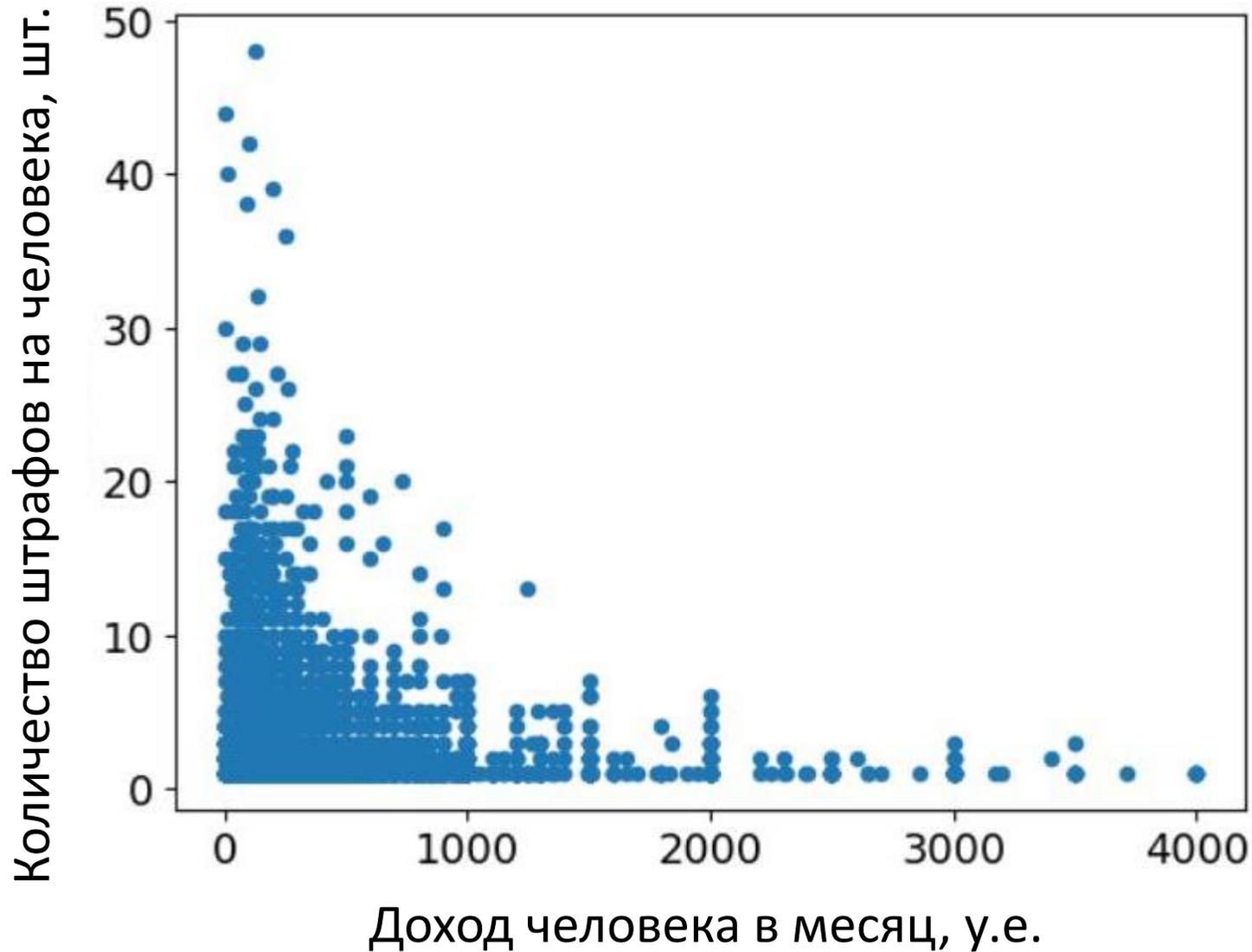
по ВОЗРАСТУ

Критерий Краскела-Уоллиса для людей младше
50 лет:
p-значение: $4.01e-38$

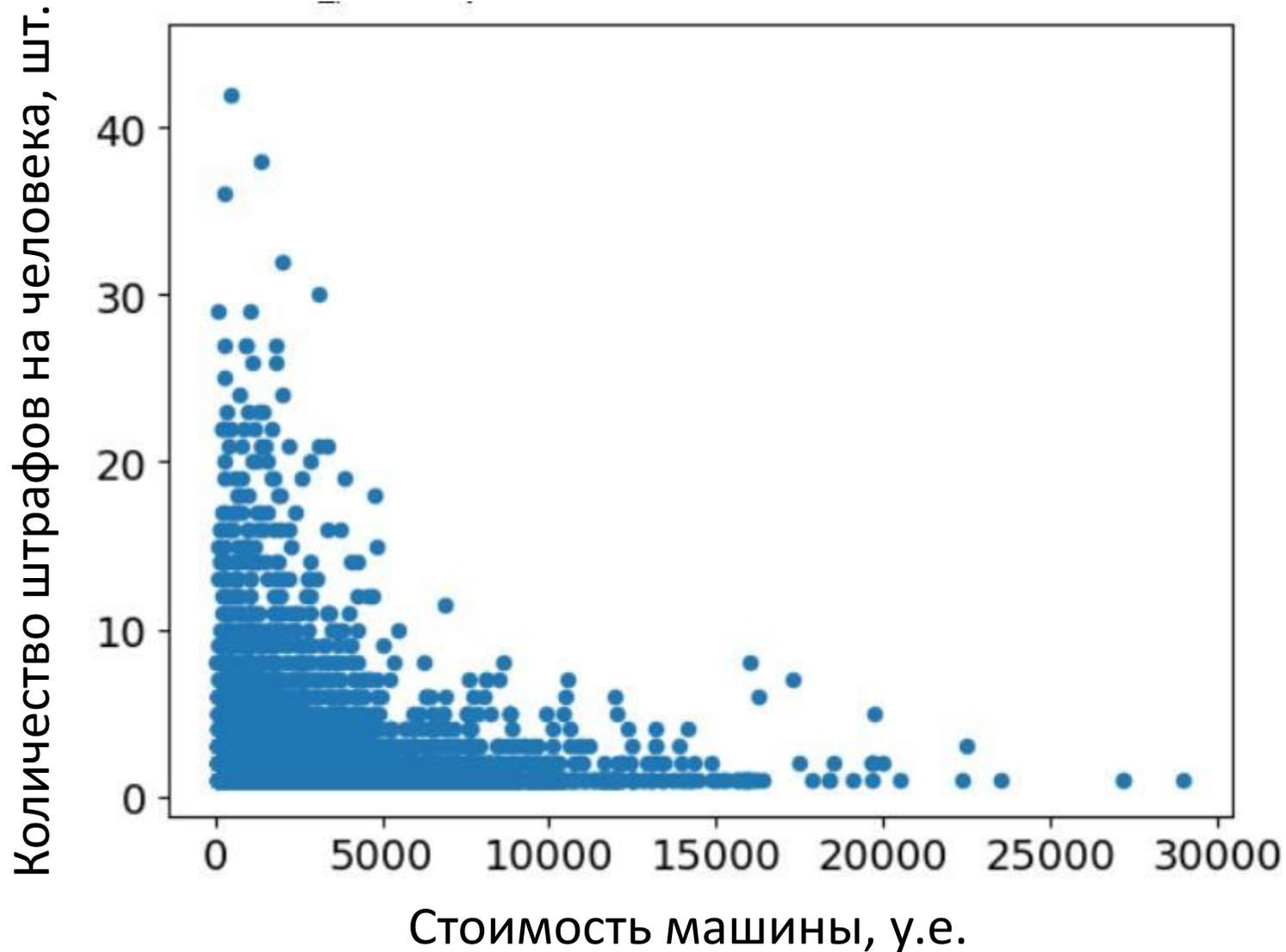
Критерий Краскела-Уоллиса для людей старше
50 лет:
p-значение: $7.18e-07$

Мат. модель устойчива

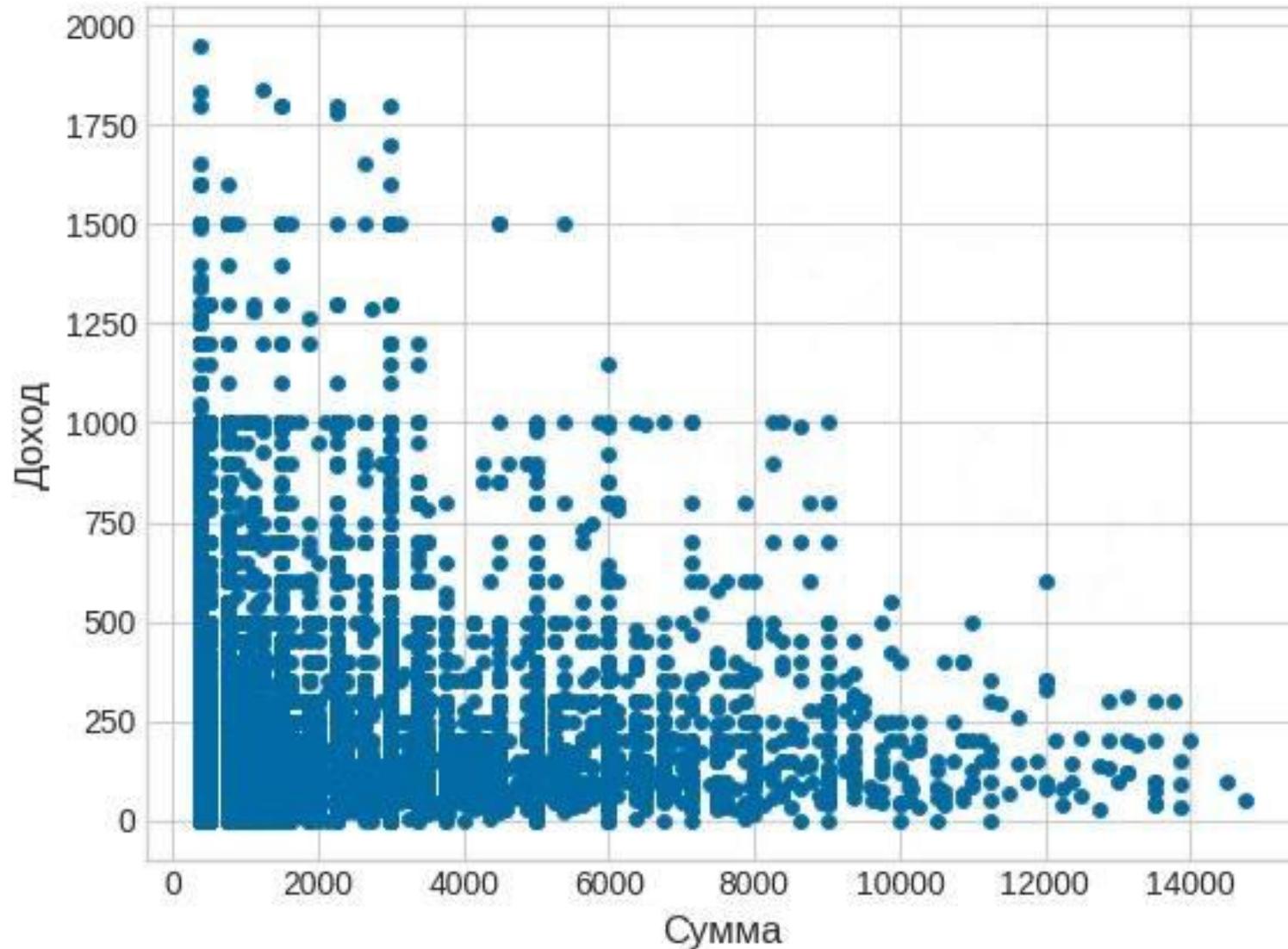
Зависимость количества штрафов от дохода



Зависимость количества штрафов от стоимости машины



Зависимость суммы штрафов от дохода



Интерпретация **НО**



Нарушители с высоким доходом имеют меньшее количество штрафов по сравнению с нарушителями из других доходных групп.

- ✓ Наша гипотеза подтвердилась
- ✓ Чем больше у человека доход, тем меньше штрафов он имеет по сравнению с нарушителями из других доходных групп.

Интерпретация **H1**



Нарушители с высоким доходом имеют **не** меньшее количество штрафов по сравнению с нарушителями из других доходных групп.

- ✓ Наша гипотеза **не** подтвердилась
- ✓ Чем больше у человека доход, тем меньше штрафов он имеет по сравнению с нарушителями из других доходных групп.

Ограничения



Обобщаемость
результатов



Нет информации о людях,
не получавших штрафы



Сложность фиксации
нарушений



Информация за короткий
временной промежуток

Перспективы

 Смотреть в разрезе

 Долгосрочная
динамика

 Обобщение

Применение

- ☞ Предложение инвестиций людям, любящим риск
- ☞ Примерное определение дохода клиента для кредитных предложений

Заинтересованы в проекте

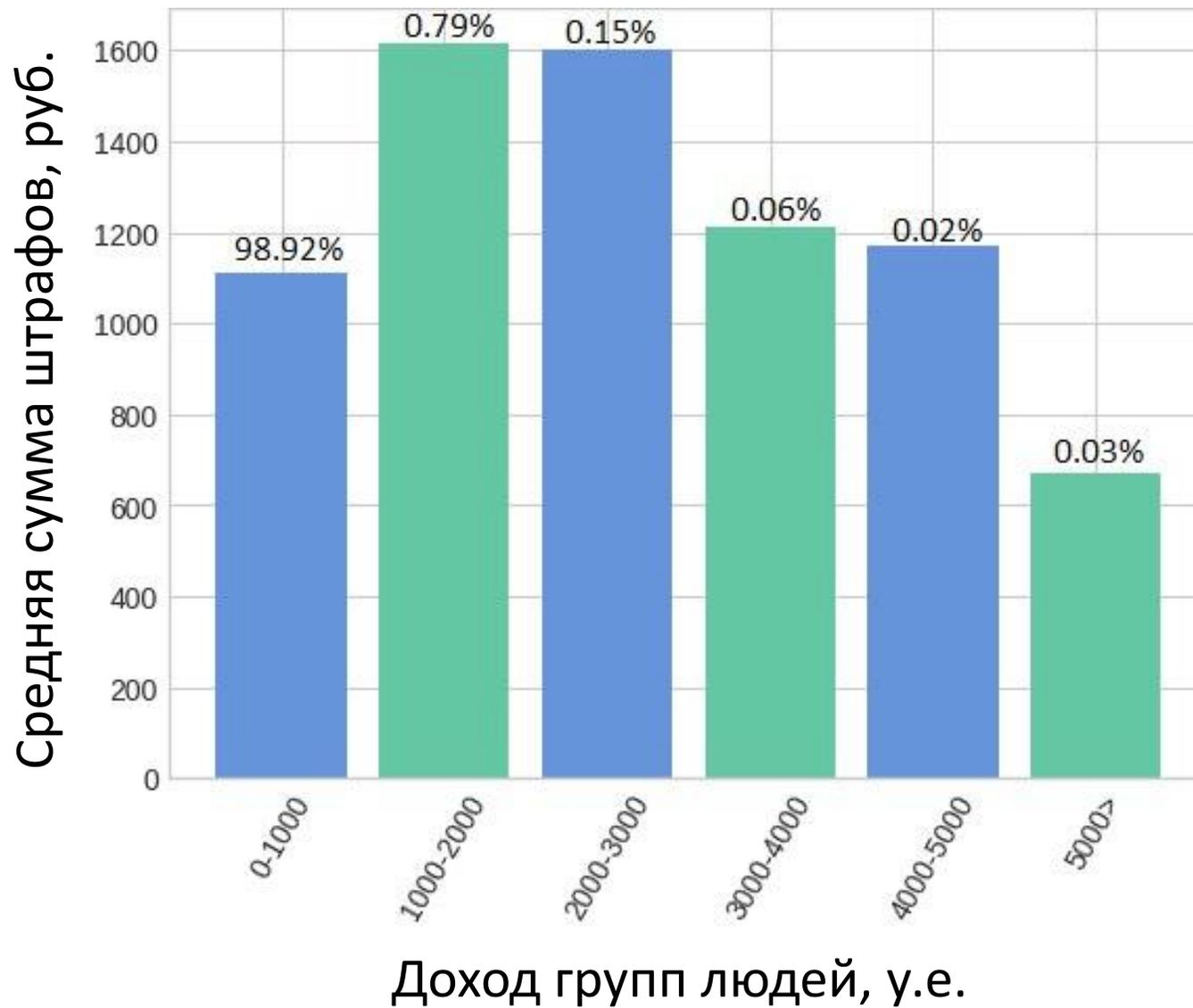
- ☞ Приложения, в которых доступна оплата штрафов
 - ☞ Органы местного самоуправления
 - ☞ Органы местного самоуправления
- гибдд

Policy implication

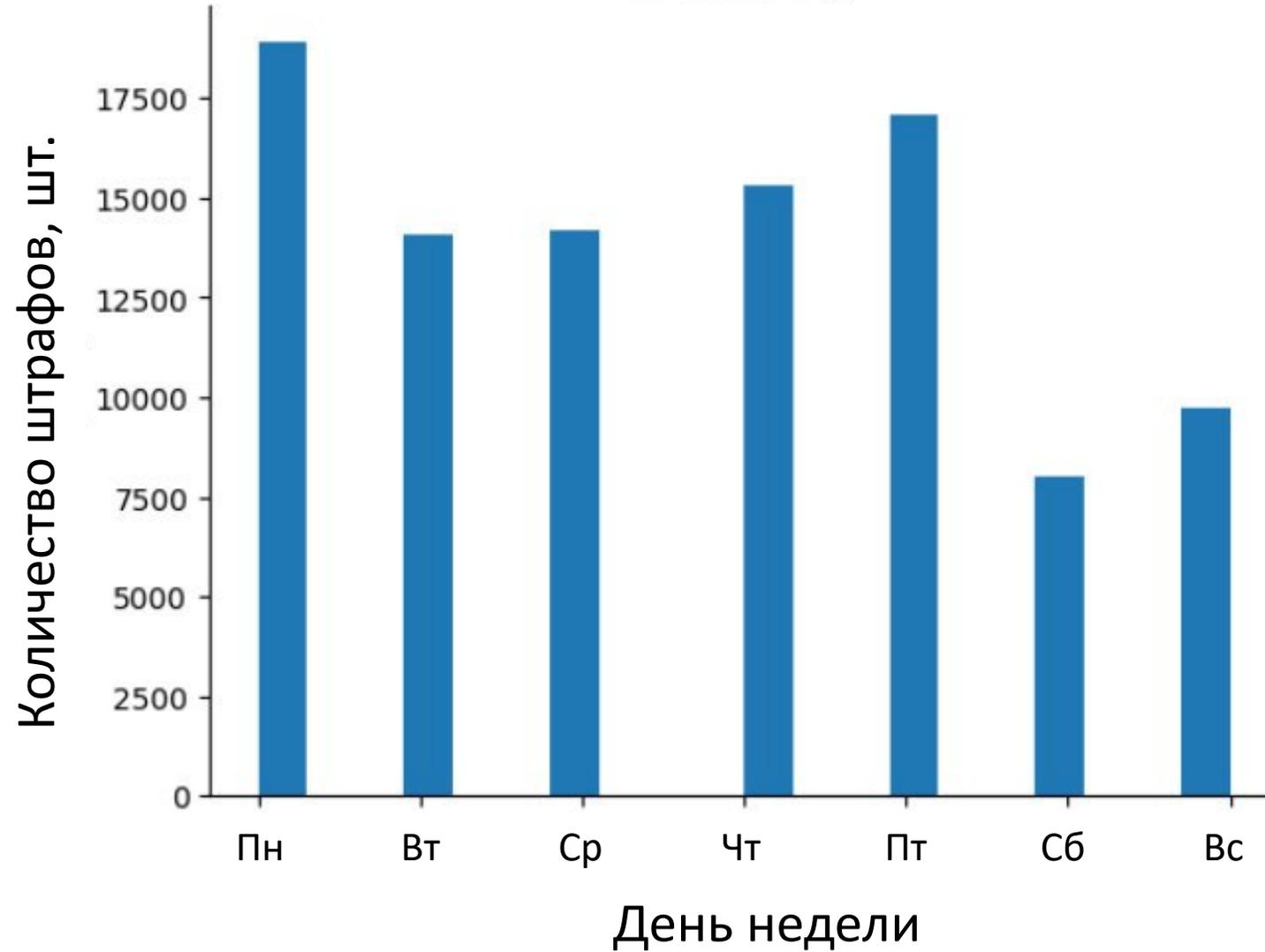
- 🏠 Введение прогрессивной шкалы штрафования в зависимости от количества нарушений за месяц
- 🏠 Кампании, нацеленные на конкретные группы с низкими доходами для улучшения безопасности на дорогах

Подвал

Зависимость суммы штрафов от дохода людей



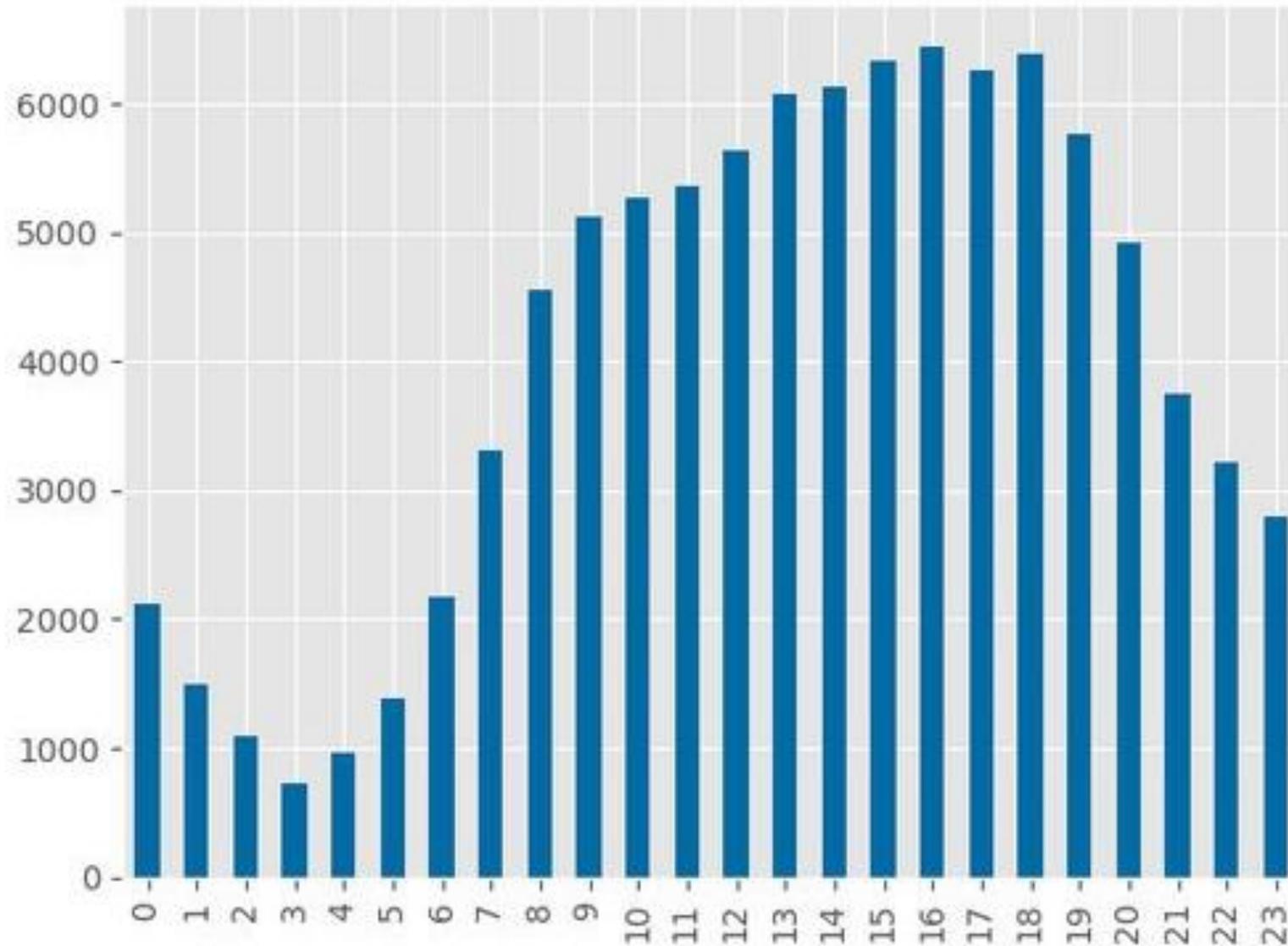
Зависимость количества штрафов от дня недели



Корреляционная матрица



Зависимость количества штрафов от времени суток



Мат. модель

ПО ПОЛУ

Критерий Краскела-Уоллиса для женщин:

Статистика критерия: 72.19

p-значение: $1.45e-15$

Критерий Краскела-Уоллиса для мужчин:

Статистика критерия: 115.32

p-значение: $7.87e-25$

ПО ВОЗРАСТУ

Критерий Краскела-Уоллиса для людей младше
50 лет:

Статистика критерия: 176.95

p-значение: $4.01e-38$

Критерий Краскела-Уоллиса для людей старше
50 лет:

Статистика критерия: 31.35

p-значение: $7.18e-07$

Мат. модель устойчива

