

DANO



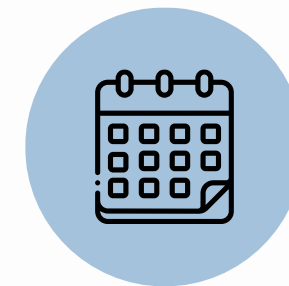
**Сравнить нельзя
купить**

КОМАНДА "ЕГЭ НЕ ВАРИК"
ЕСТЕХИН НИКИТА,
НИКОЛАЕВ ЕГОР,
ГЕРАСИМОВ ДАНИИЛ,
НЕКРАСОВА ВЕРОНИКА

ДАТАСЕТ

Данные о взаимодействиях пользователей с цифровыми каналами «Спортмастера» на этапах воронки онлайн-продаж.

7 миллионов записей



Период сбора данных: с
01.06.2024 по 21.06.2024

**Более 2 миллионов
пользователей**




Типы устройств: desktop,
Android, IOS

Более 300 тысяч заказов



Характеристики товаров:
категория, гендер, возраст и
варианты получения

Обработка технических ошибок

 Очищаем от записей когда были куплены товары, у которых **отсутствовали все типы доставки**

1817 записей

 Удаляем записи, где **дублируется event_id**

34 записи

 Удаляем записи с **пропусками**

11 записей

*Запись - действие пользователя на платформе, добавленное в датасет

Обработка выбросов

Причины

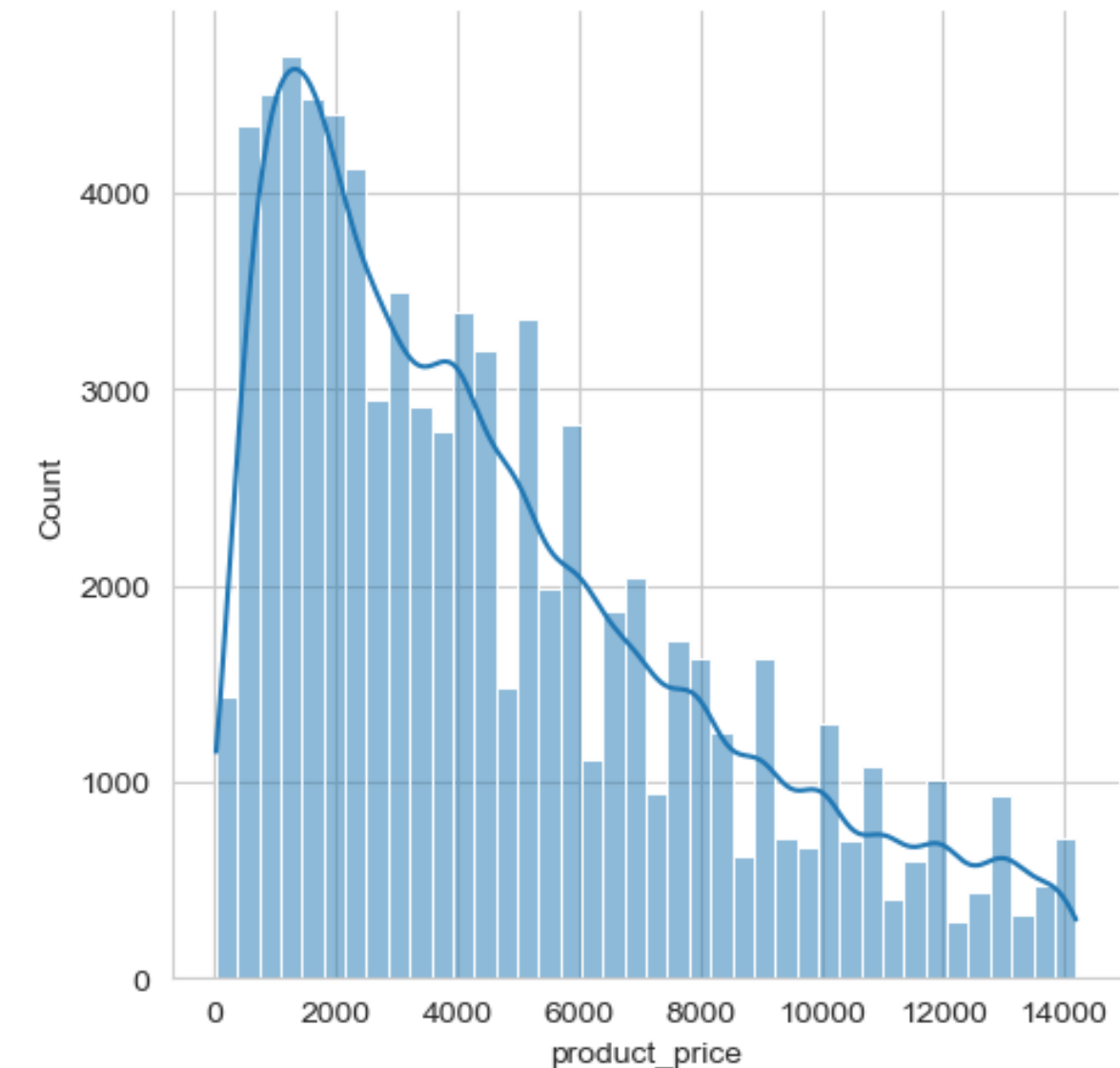
- Среднее арифметическое, которое мы будем использовать, **неустойчиво** к выбросам
- В исследовании мы рассматриваем только **основную часть** товаров - без очень дорогих продуктов

Очистка выбросов

Удаляем выбросы в цене товаров методом межквартильного размаха

**574 655 записей
(8 %)**

Распределение цен после чистки



Подготовка данных

Введение новых переменных / понятий

is_discount - наличие скидки на товар (0, 1)

pct_discount - **процент** скидки на товар

price_category - ценовая категория, деление через квантили на **"Low", "Mid", "High", "Top"**

discount_category - категория скидки, без скидки - **"No"**,
оставшееся через квантили на **"Small", "Medium", "Large", "Huge"**

Объединение android и ios

ПРИЧИНА

Направление исследования - анализ различий
десктопа и мобильного приложения

ЗАДАЧА

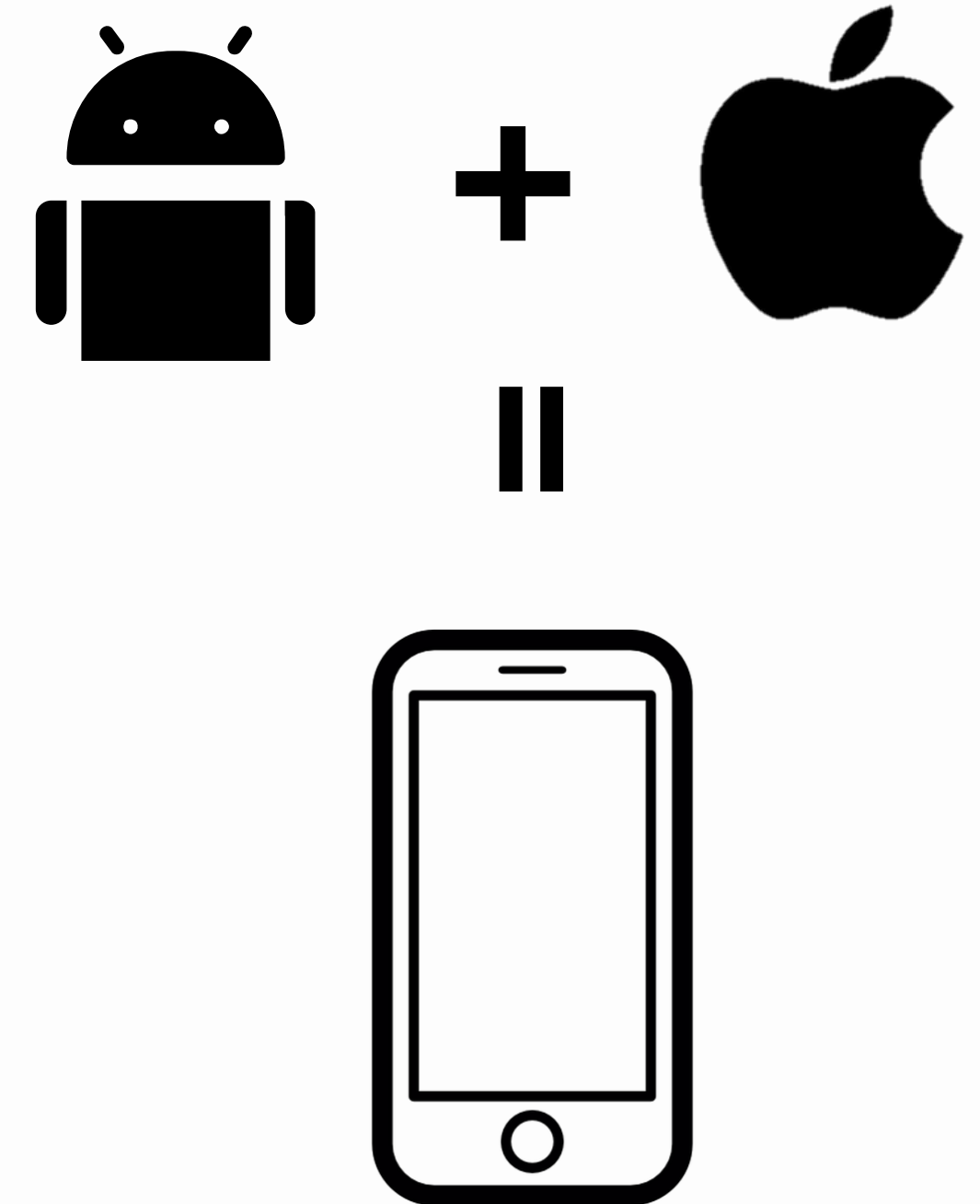
Повышение **достоверности** статистических тестов,
упрощение проведения **анализа**

desktop

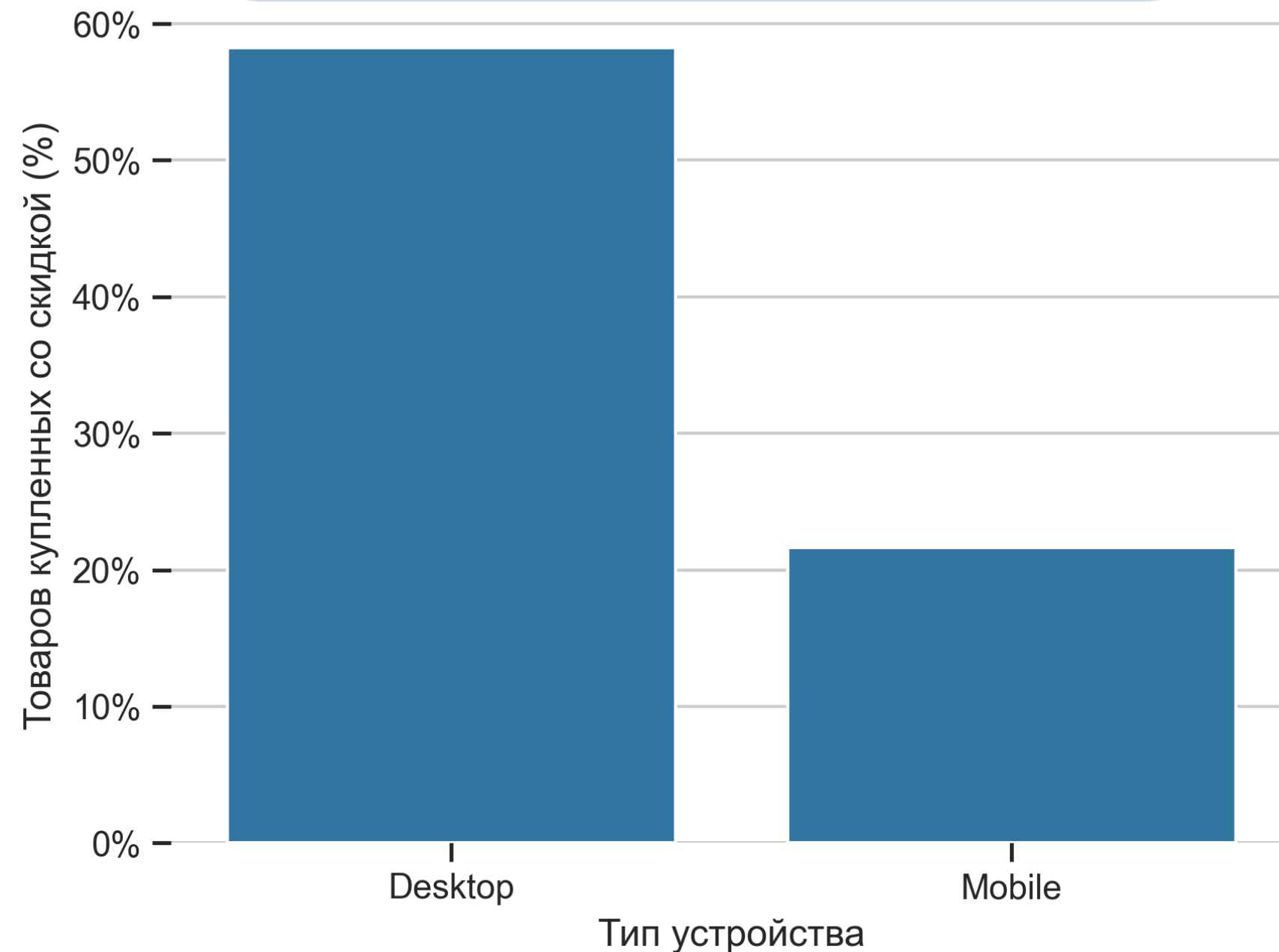
mobile

Десктопная версия сайта

Мобильно приложение на
Android и IOS

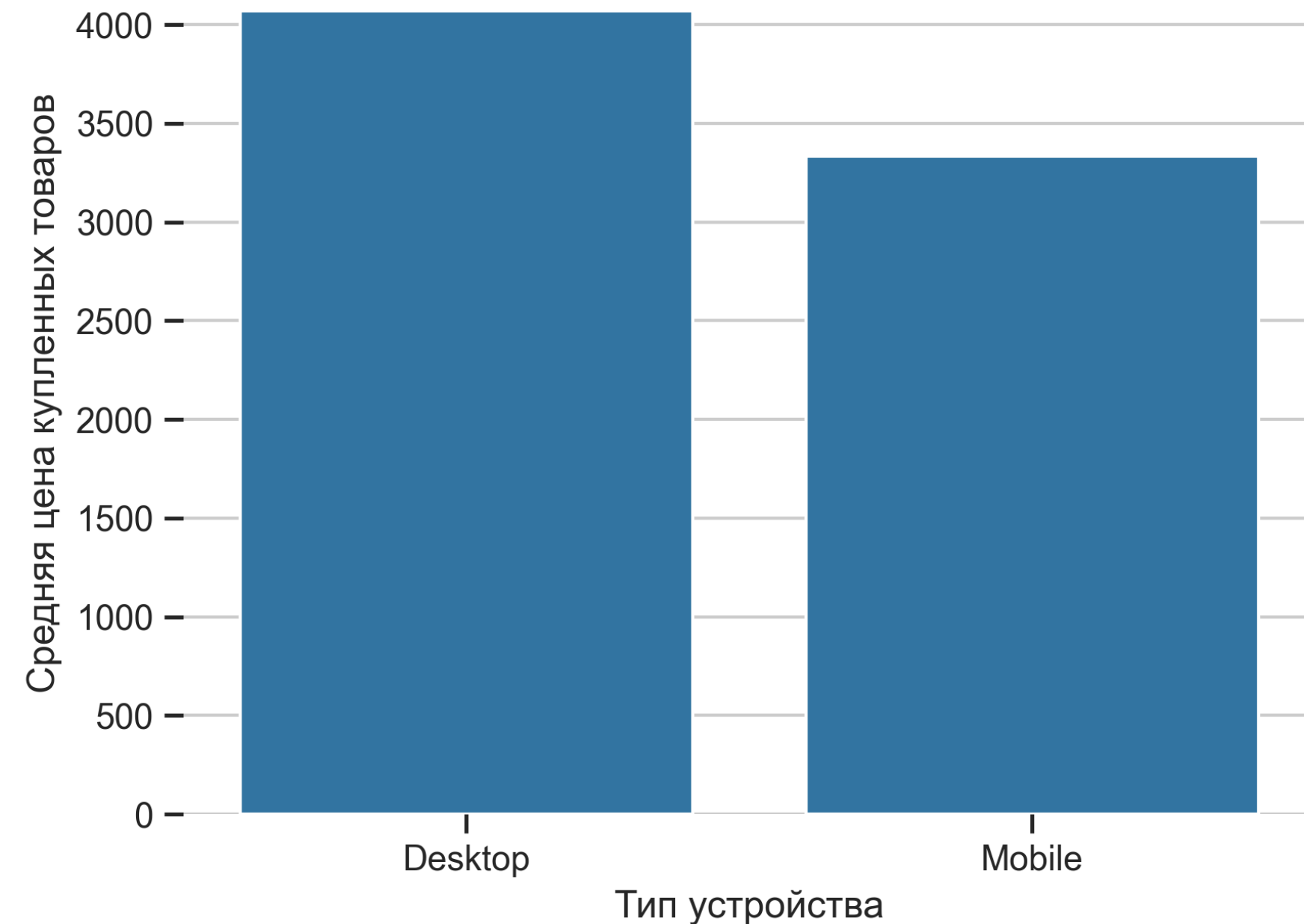


Процент товаров купленных со скидкой по типам устройств



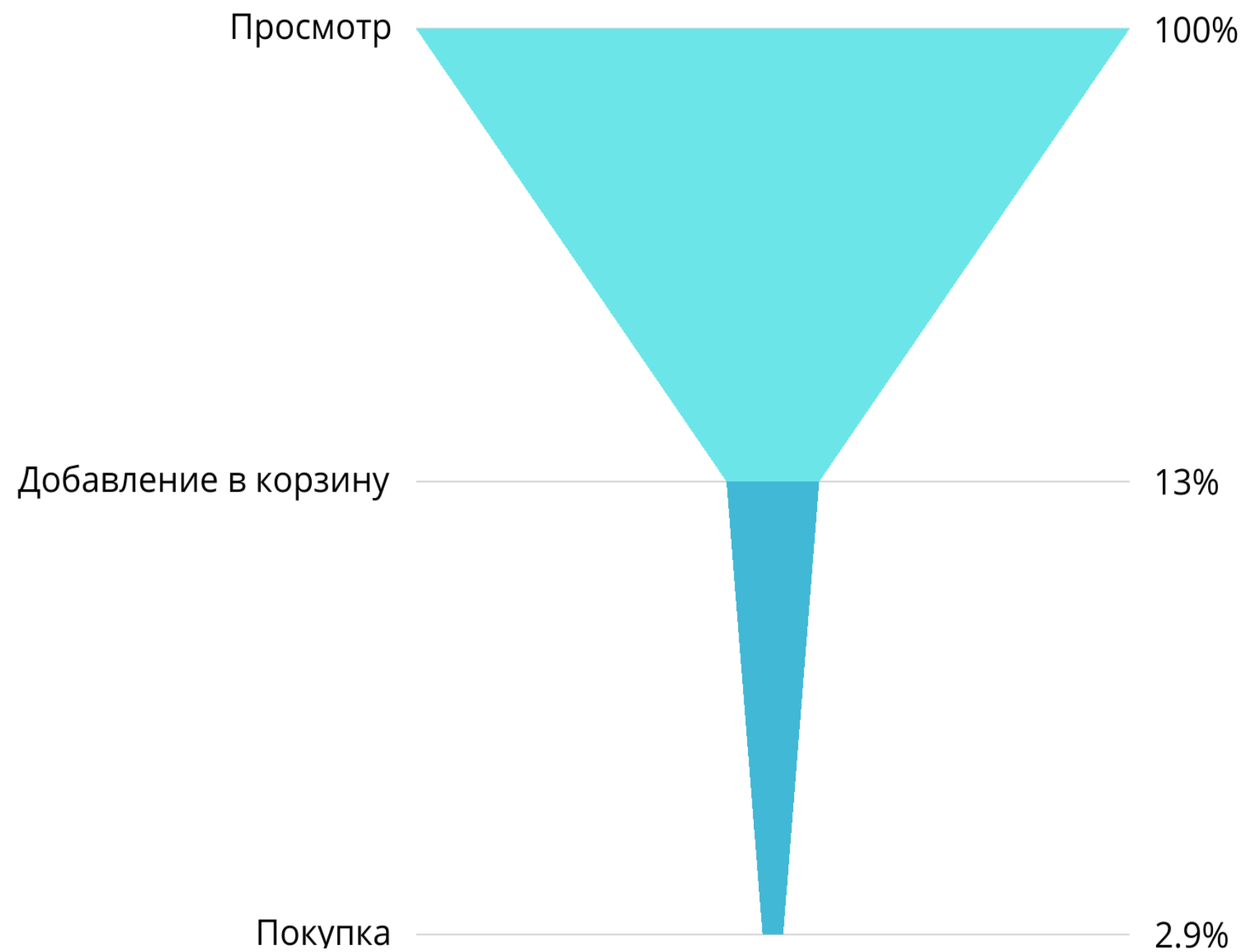
Процент товаров, купленных со скидками, по типам устройств на **десктопе выше**

Средняя цена купленного товара по типам устройств

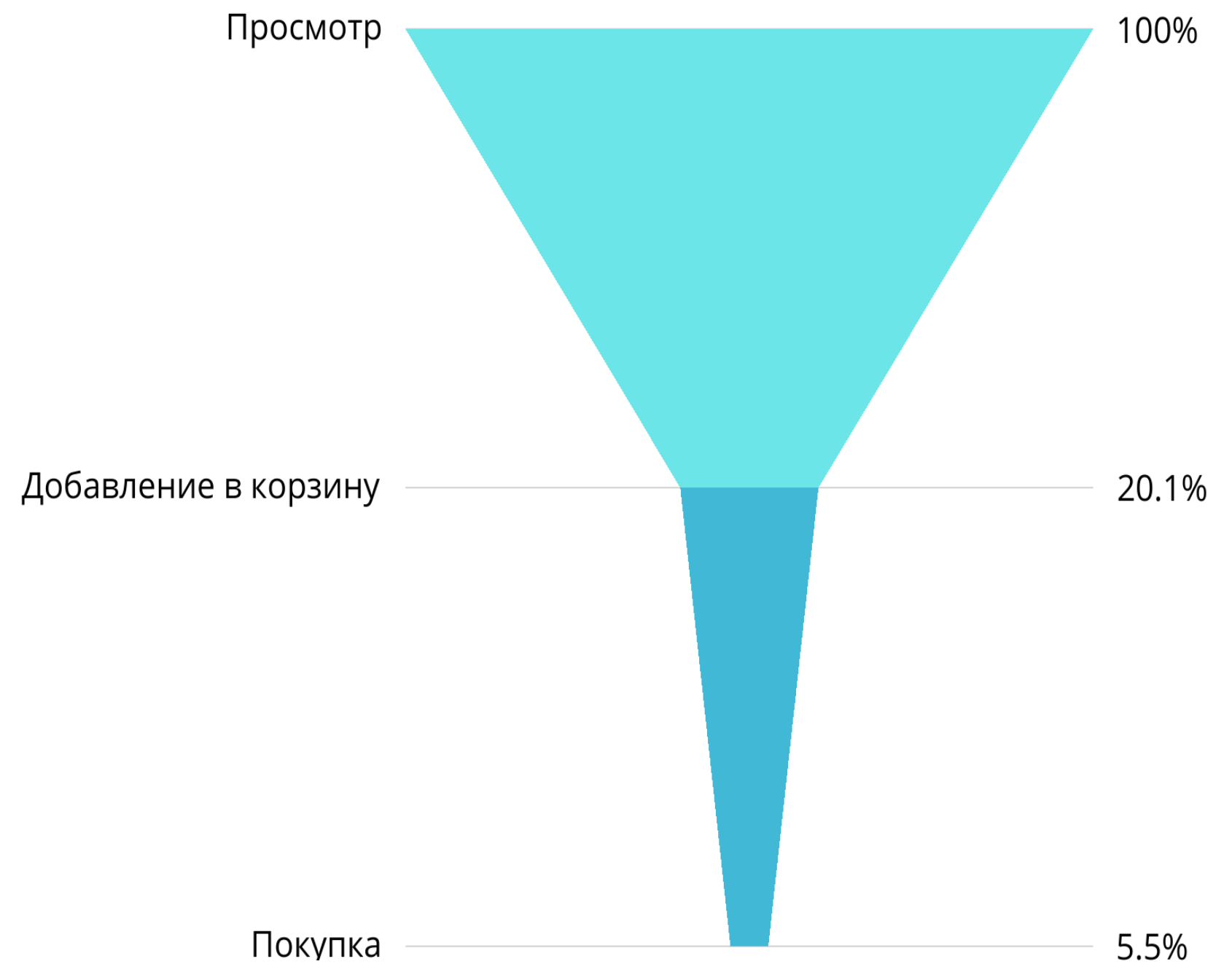


Средняя цена по типам устройства на **десктопе выше**

Конверсии этапов воронки для десктопа



Конверсии этапов воронки для приложения



ИССЛЕДОВАТЕЛЬСКИЙ ВОПРОС И ГИПОТЕЗА

Исследовательский вопрос

Какие факторы влияют на
конверсию прохождения
этапов воронки
интернет-магазина ?



Гипотеза

Конверсия прохождения
этапов воронки **на**
десктопе ниже, чем в
мобильной версии
приложения.

МЕХАНИЗМ ГИПОТЕЗЫ



Покупаются **дорогие**
товары



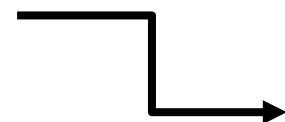
Желание **найти**
наилучший вариант



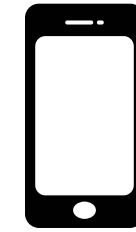
Покупается **больше**
товаров **со скидками**



Тратится **больше**
времени на анализ



Конверсии на
компьютерах
ниже



Покупаются **недорогие**
товары



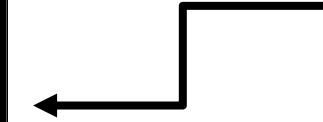
Не возникает
сомнений в покупке



Реже покупаются
товары **со скидками**



Тратится **меньше**
времени на анализ



Конверсии на
телефонах
выше

Статистическое подтверждение гипотезы

H0

В генеральной совокупности **нет** статистических **различий** конверсий по типам устройств

H1

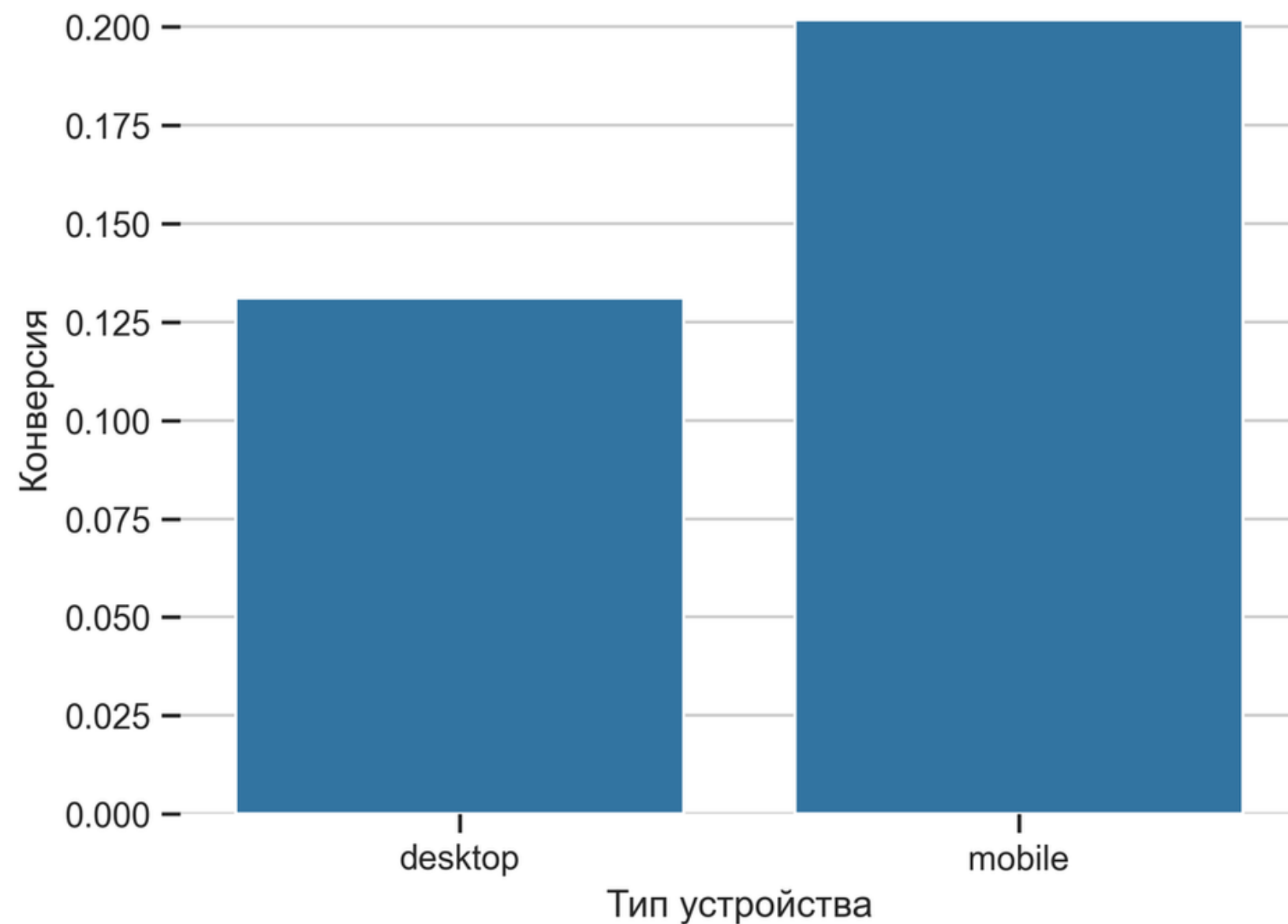
В генеральной совокупности **есть** статистическое **различие** конверсий по типам устройств

5%
Уровень доверия

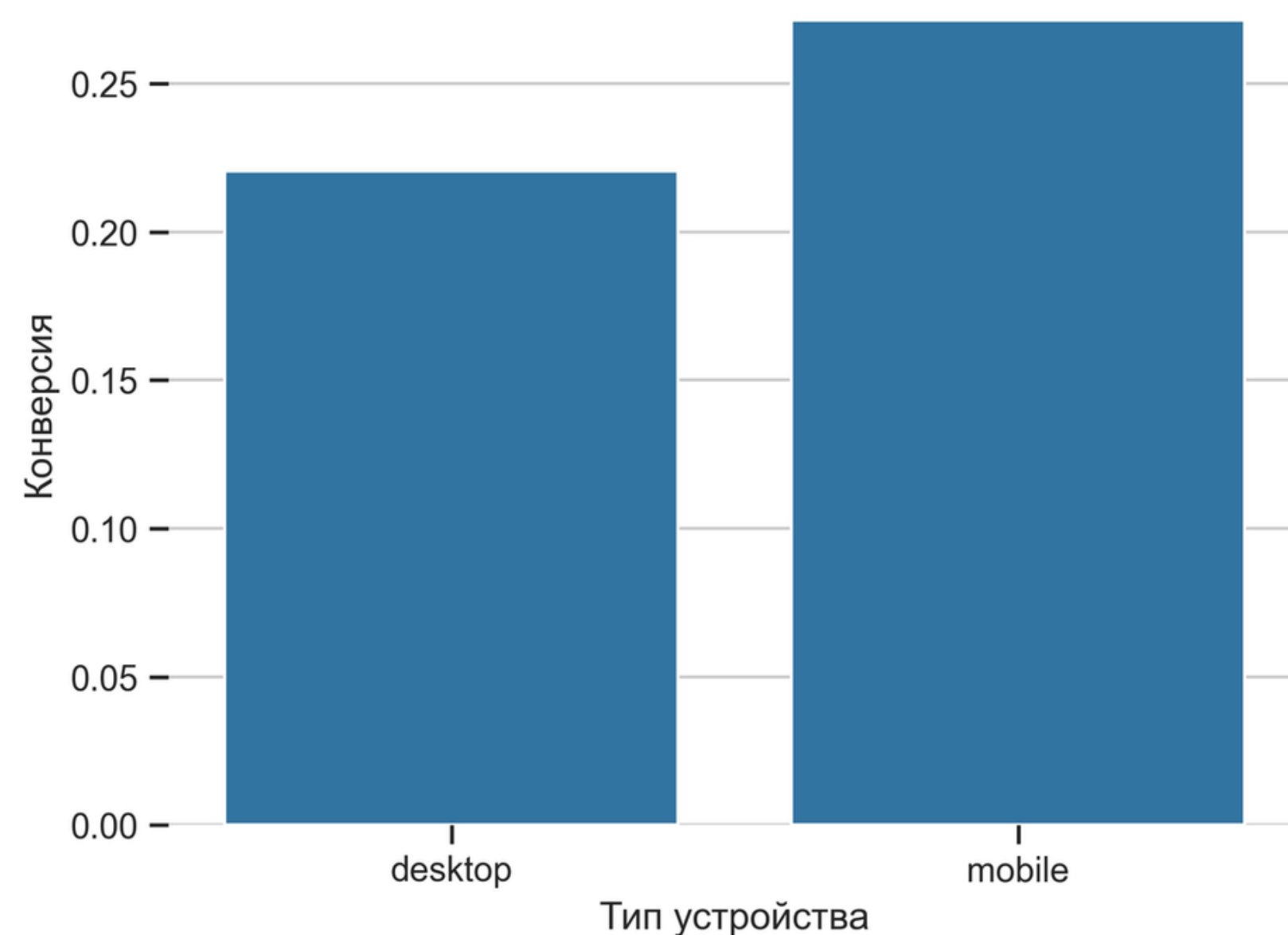
Математическая модель

Конверсии прохождения этапов воронки по типам устройств

Из просмотра в добавление в корзину



Из корзины в покупку



Математическая модель

	Просмотр	Добавление в корзину
Desktop	679 278	89 078
Mobile	5 744 339	1 158 388

Используем Хи-квадрат Пирсона два раза для *двух этапов воронки*:

- 1.Просмотр - Добавление в корзину
- 2.Корзина - Покупка

	Добавление в корзину	Покупка
Desktop	89 078	19 637
Mobile	1 158 388	314 210

Почему именно хи-квадрат Пирсона ?

- 1.**Таблица сопряжённости** с категориальными данными
- 2.Необходимо определить статистические различия между **тремя группами**

Статистическое подтверждение

Из просмотра в
добавление в корзину

chi2 13666

p-value 0

Из корзины в
оформление заказа

chi2 651

p-value 0

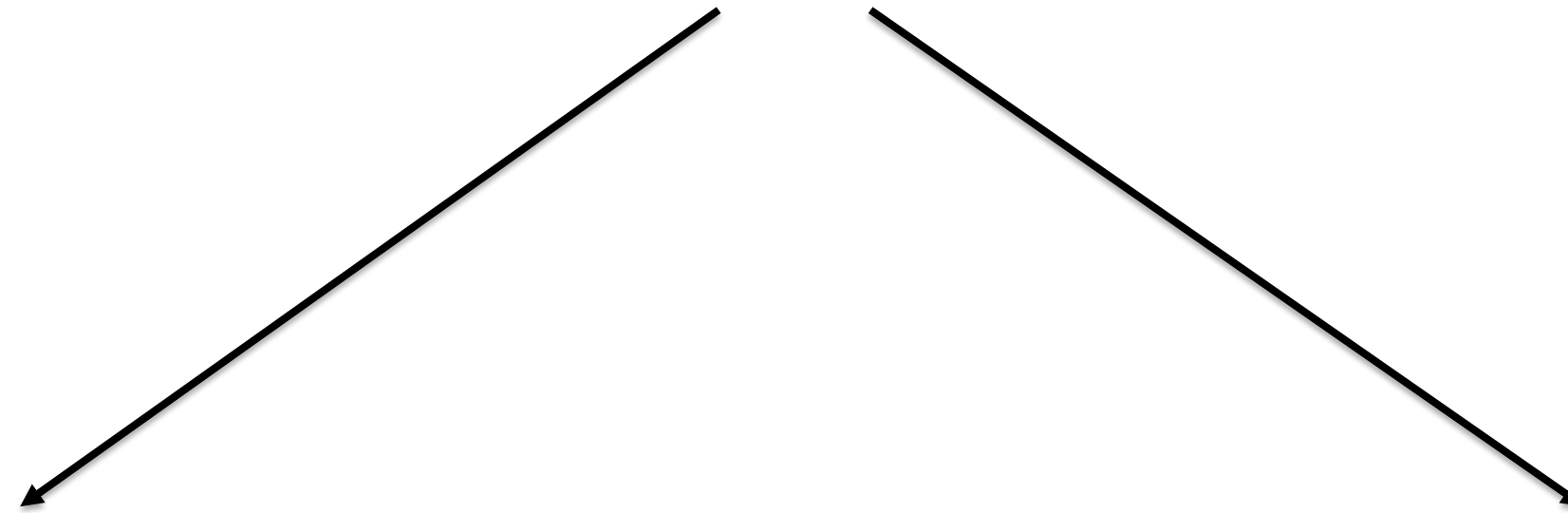
$p\text{-value} < 0.05$ в обоих случаях, **нулевая гипотеза отвергается**, конверсии на обоих этапах зависят от типа устройства

Статистическое подтверждение механизма

Часть механизма	Статистический тест	P-value
Средняя цена купленных товаров на десктопе выше, чем в приложении	t-критерий Уэлча	0
Средний процент скидки купленных товаров на десктопе выше	t-критерий Уэлча	0
Процент товаров, купленных со скидкой на десктопе выше, чем в приложении	Хи-квадрат	0

*стат. подтверждение в приложении стр. 1-6

Рассмотрим другие факторы, которые могут
влиять на конверсию по типам устройств



На разных типах устройств покупают
разные категории товаров (не
рассмотренные ранее)

На конверсию влияют
особенности устройств



Проверка **различий** в **количестве** покупаемых с **разных типов устройств** товаров

Категория	p-value	Метод
Тип товара	> 0.05	Хи-квадрат
Возраст товара	> 0.05	Хи-квадрат
Гендер товара	> 0.05	Хи-квадрат
И др.

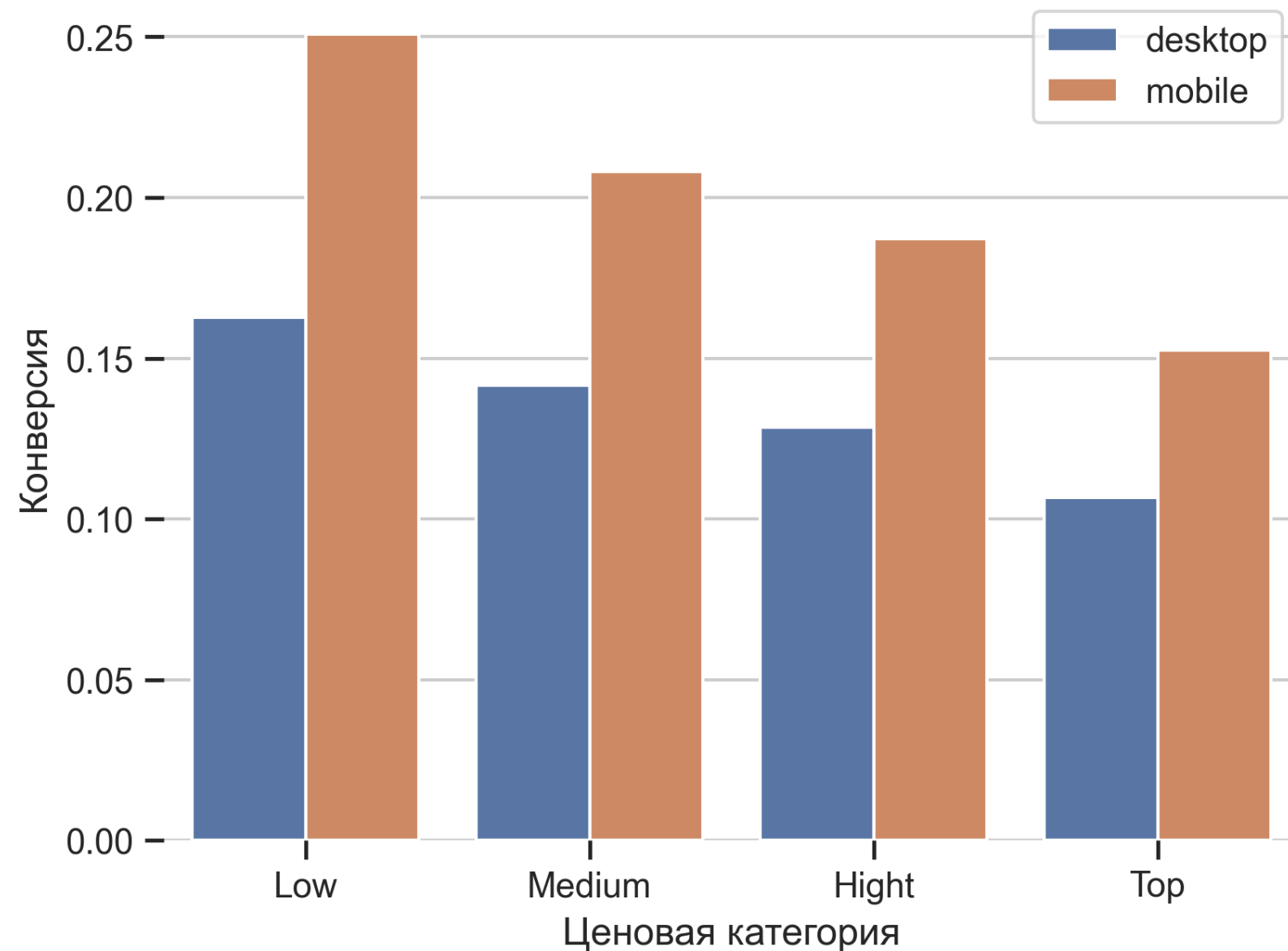
Значимые различия

- Категория цены
- Категория скидки

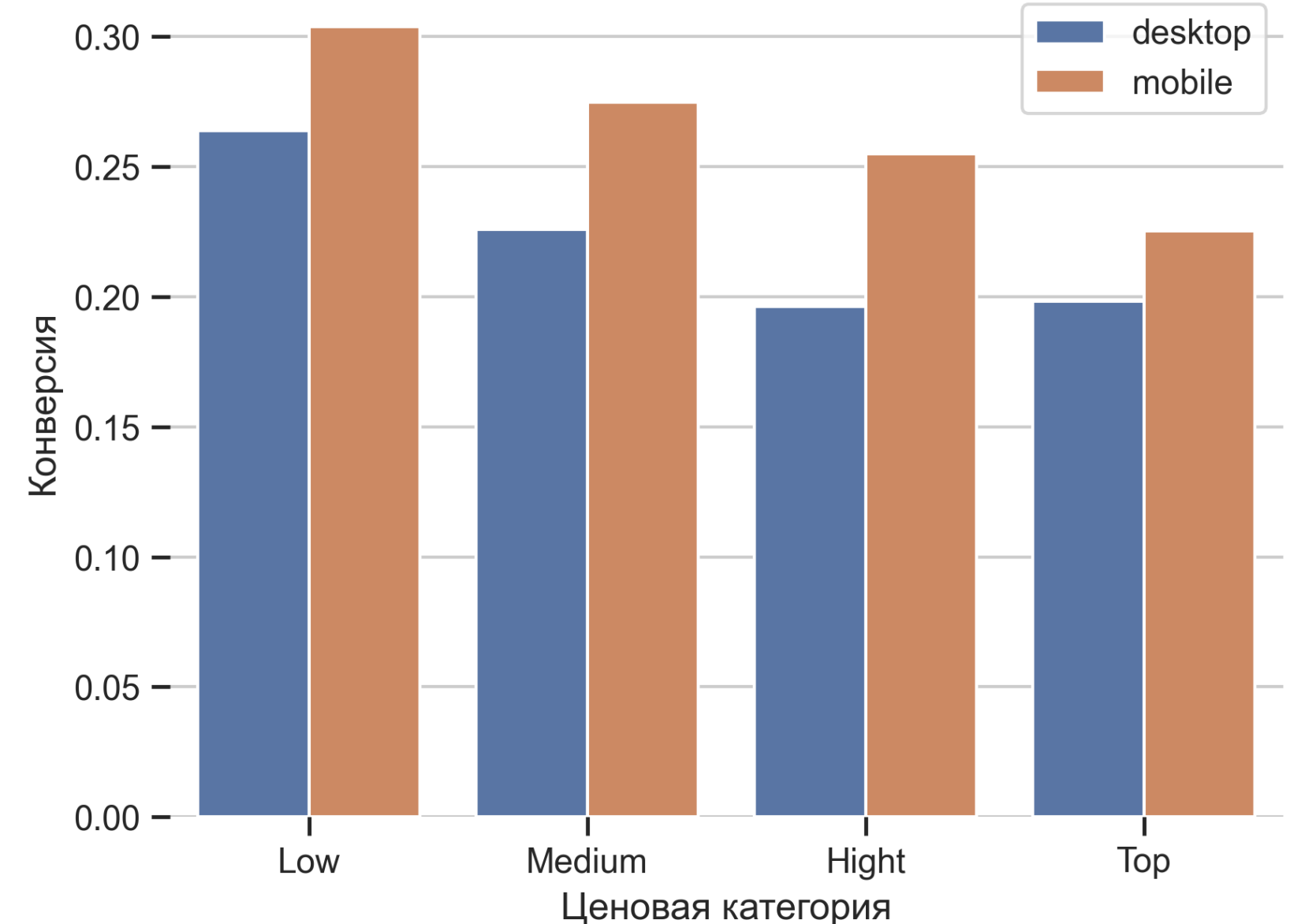
*стат. обоснование в приложении стр. 7-15

Конверсии по типам устройств в разрезе категории цены

Из просмотра в добавление в корзину



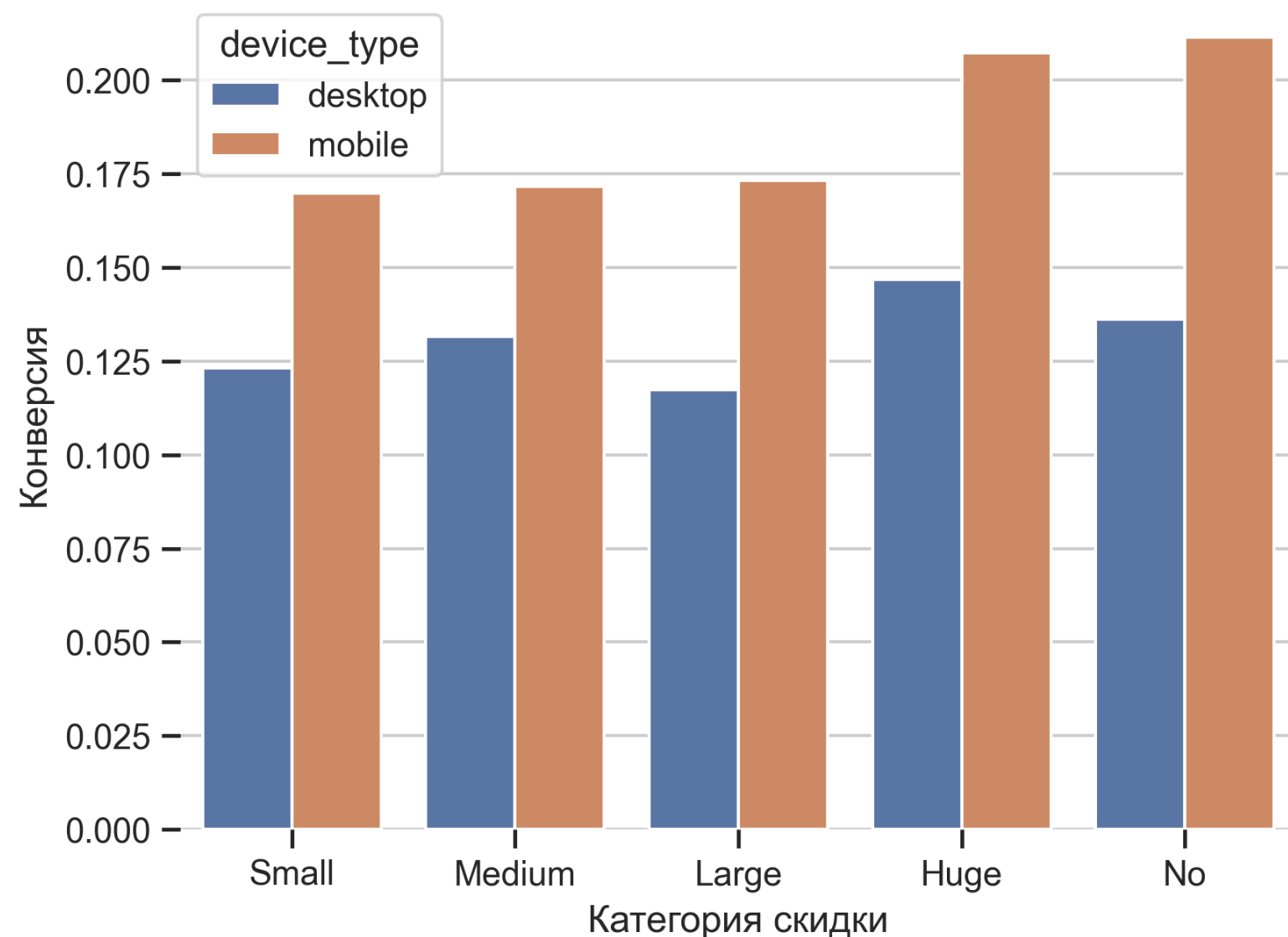
Из корзины в покупку



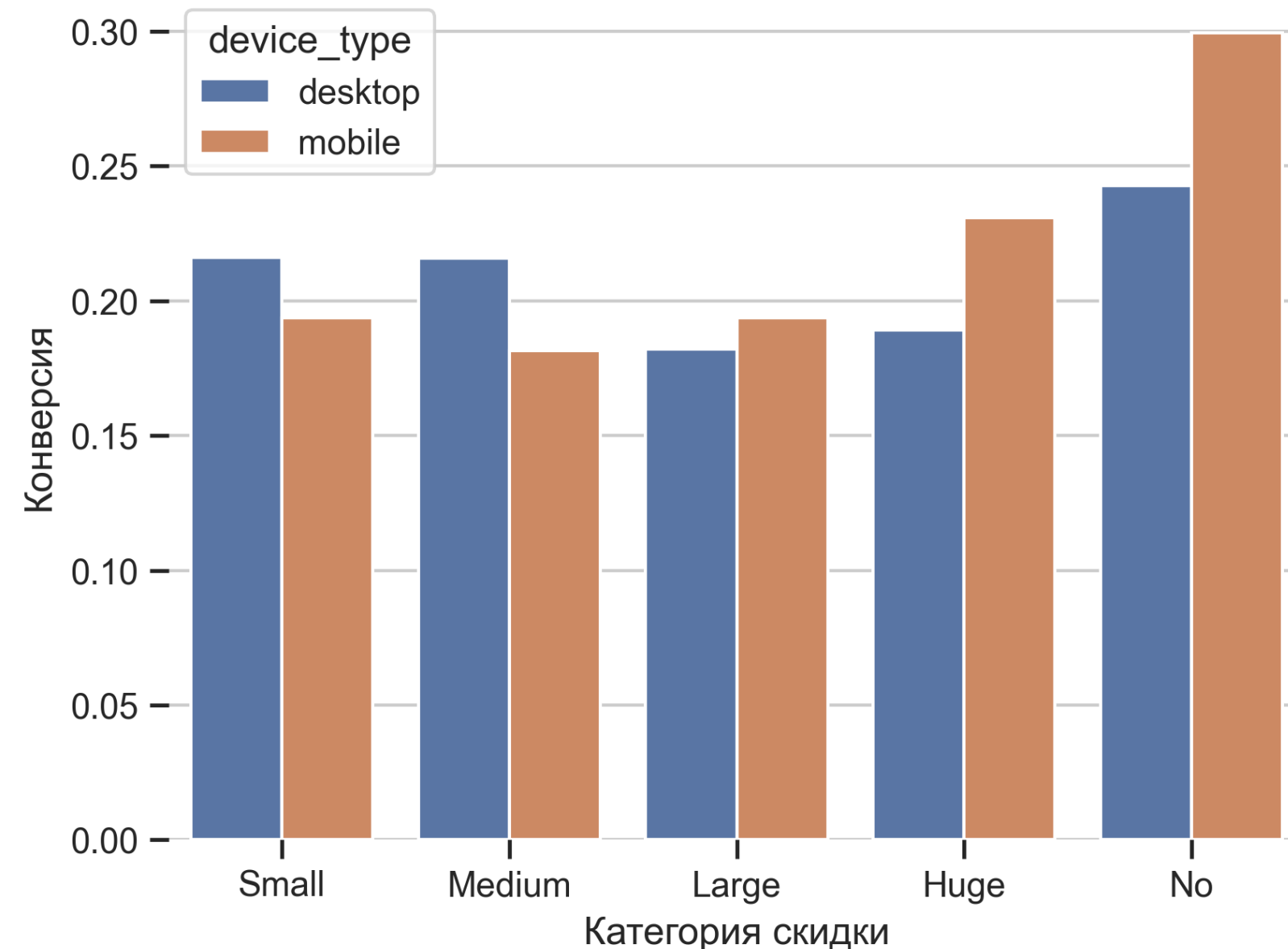
Во всех ценовых категориях при проверке различия конверсий по типам устройств
 $p_value \approx 0$

Конверсии по типам устройств в разрезе категории скидки

Из просмотра в добавление в корзину



Из корзины в покупку



Во всех группах, **кроме категорий скидок «Small», «Medium» на втором этапе**, конверсия на десктопе ниже, а в указанных категориях выше. При проверке хи-квадратом **p-value < 0.05 для каждой группы**

*стат. обоснование в приложении стр. 12-15

Вывод об альтернативном механизме

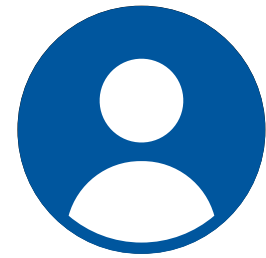
Влияние на конверсию

Подтвердилось	Не подтвердилось
<ul style="list-style-type: none">Маркеры цена и скидка товаровОсобенности платформ	<ul style="list-style-type: none">Разница покупаемых товаров

Устойчивость матмодели на остальных выборках

Выборки, которых не работает модель

Размер выборки



Авторизованные пользователи

1 475 626 пользователей



Товары с категориями скидок **“Small”**,
“Medium” (второй этап воронки)

“Small”: **89 074** записей

“Medium”: **97 954** записей

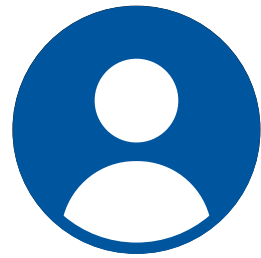


Товары категории **взрослая одежда**
премиальных брендов (второй этап
воронки)

80 787 записей

Устойчивость матмодели на остальных выборках

Выборки, которых не работает модель



Авторизованные пользователи



Товары с категориями скидок **“Small”**,
“Medium” (второй этап воронки)



Товары категории **взрослая одежда**
премиальных брендов (второй этап
воронки)

Причины поведения

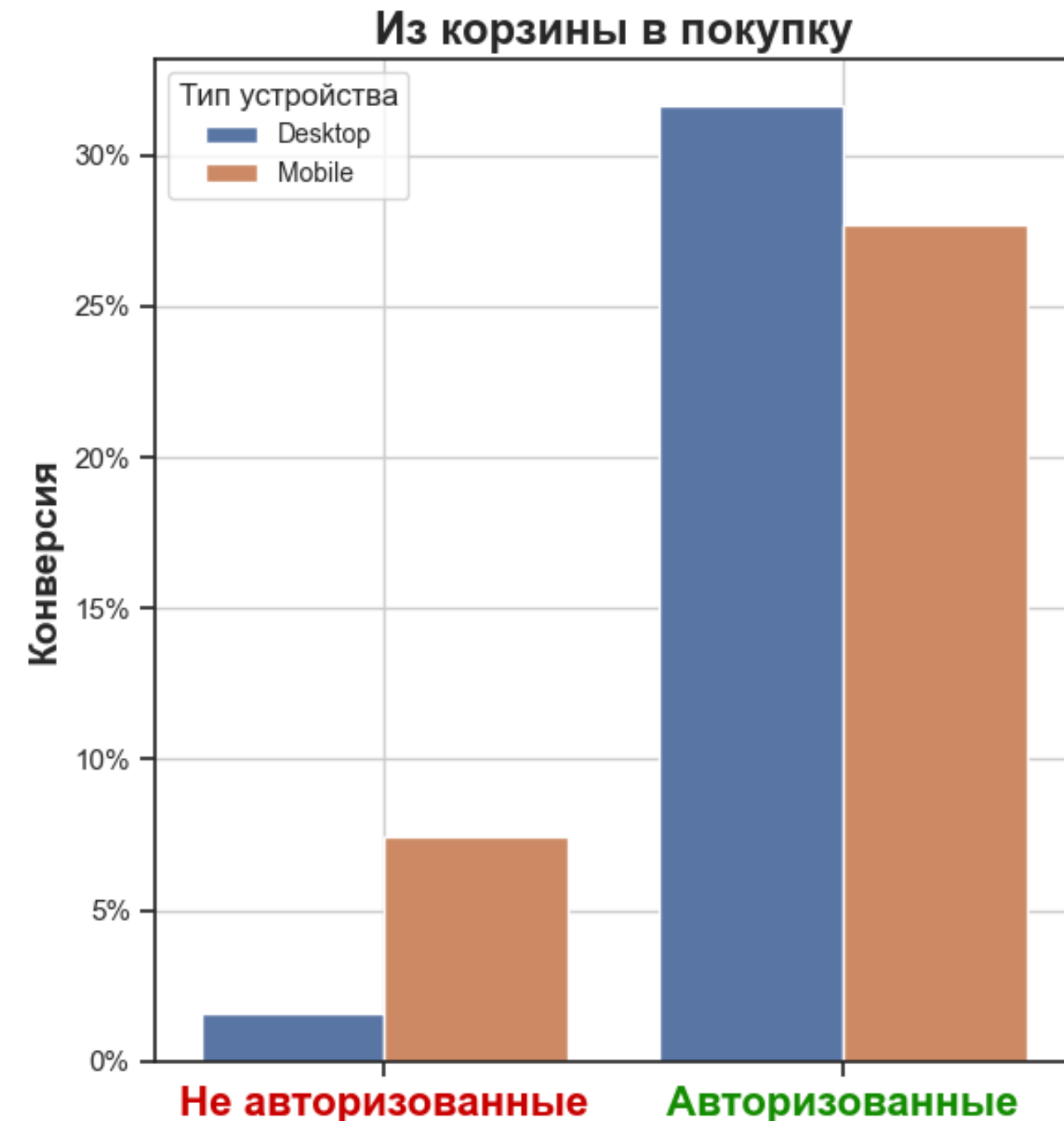
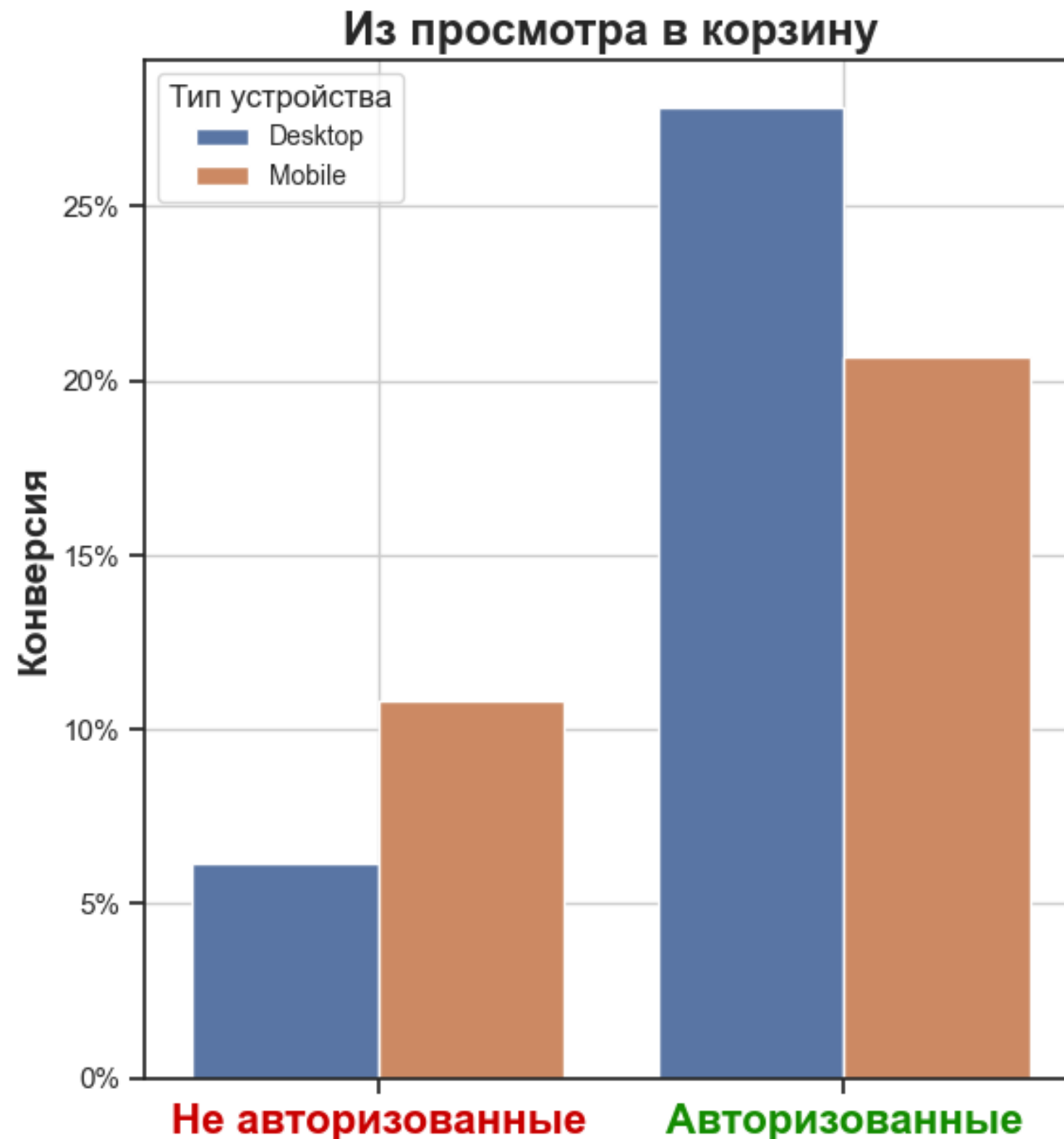


Пользователи десктопа
чувствительны к скидке. Она
повышает их конверсию

Клиенты, покупающие их,
обеспечены, поэтому
меньше сомневаются при
покупке дорогих товаров

Выборка авторизованных пользователей

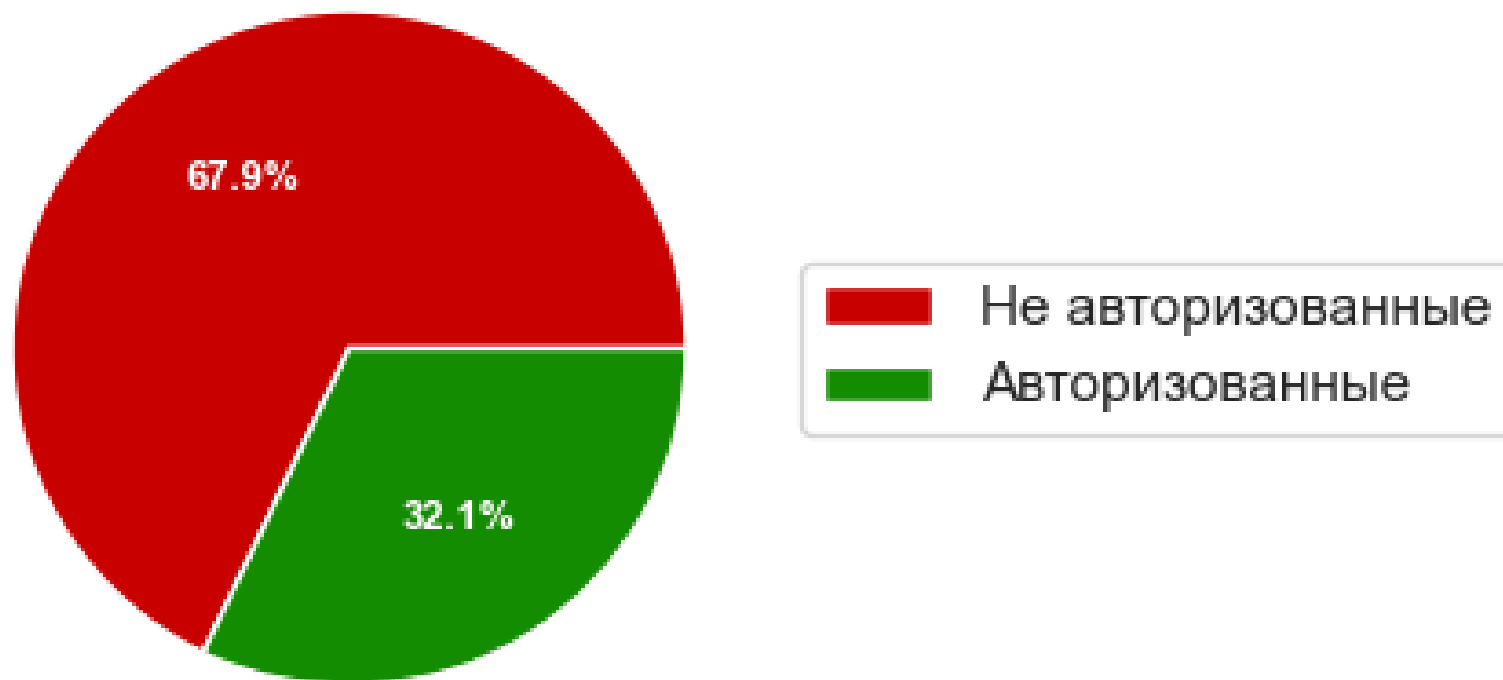
Конверсии



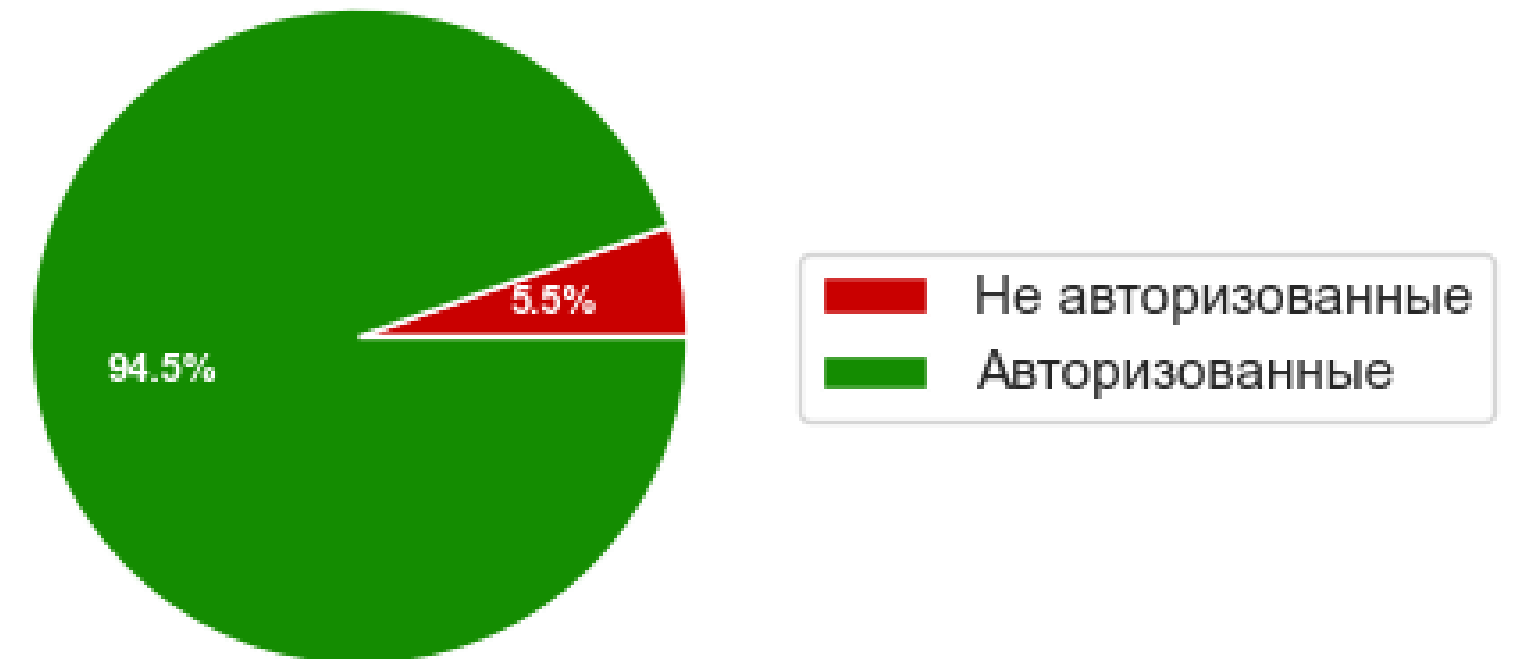
Выборка авторизованных пользователей

Доля записей **неавторизированных** пользователей на разных типах устройств

Desktop



Mobile



Очень большая доля записей неавторизованных пользователей на десктопе

<div>24</div>	<div> <div>Работа с данными</div> <div>Преданализ</div> <div>Мат модель</div> <div>Стат. подтверждение гипотезы и механизма</div> <div>Расширение механизма</div> <div>Проверка устойчивости модели</div> <div>Выводы</div> </div> <div> <div> <div></div> </div> <div> <div>Авторизация в самом начале</div> </div> </div>
<div>Этап</div> <div>Состояние пользователя</div>	<div> <div> <div> <div>Лента товаров</div> <div>Карточка</div> <div>Корзина</div> <div>Покупка</div> </div> <div> <div></div> <div></div> <div></div> <div></div> </div> </div> <div> <div> <div>Не авторизован</div> <div>Авторизован</div> </div> <div> <div>Предложение регистрации</div> </div> </div> </div>

*пояснение в приложении стр.20-22

Выводы

Гипотеза подтвердилась



Конверсия на десктопе **ниже**, чем
в приложении



Из-за разницы в **цене** и
скидке покупаемых товаров



Из-за **особенностей**
платформы

Однако

Гипотеза не подтвердилась на
выборках:



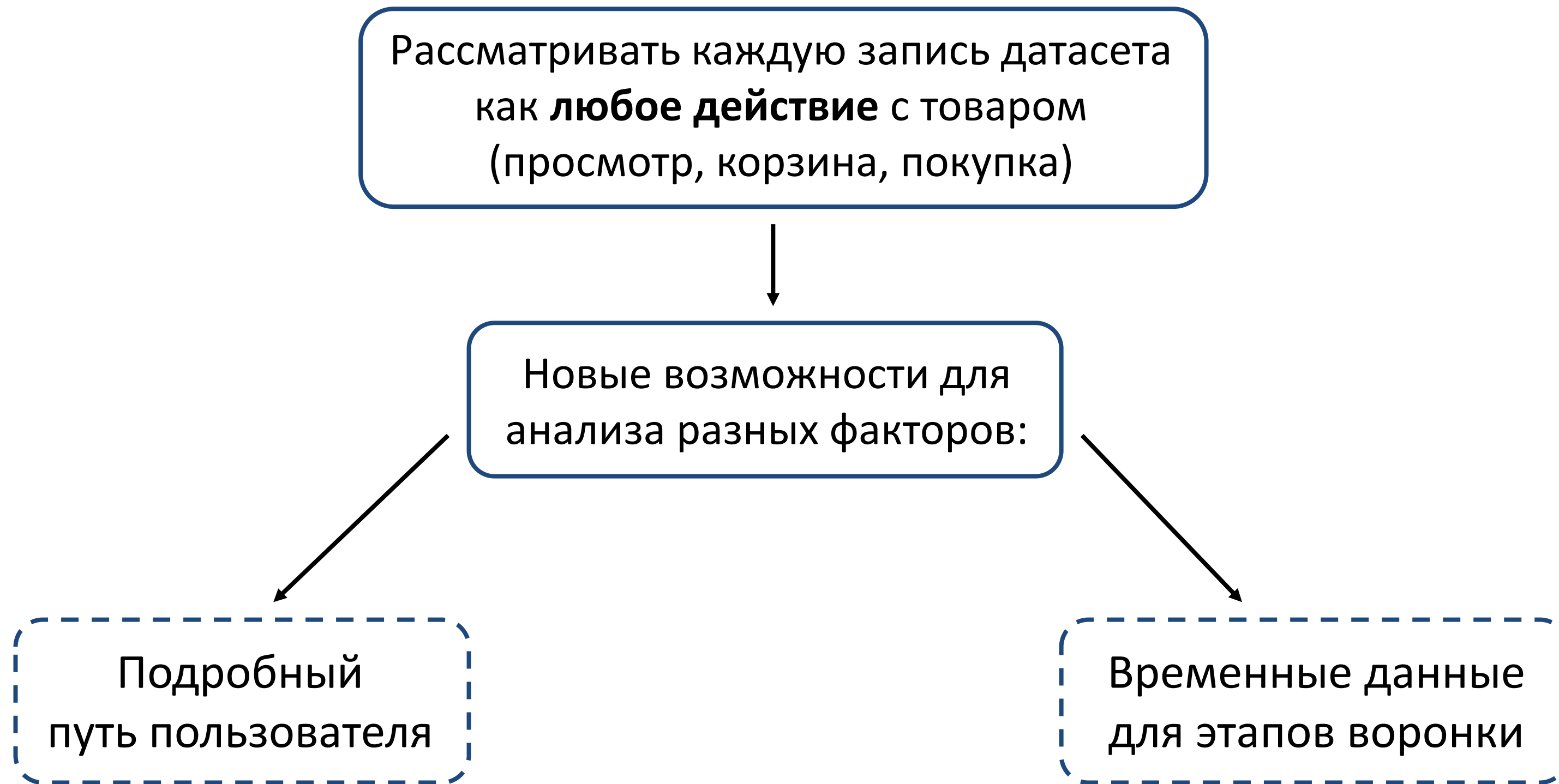
Пользователей **десктопа**,
покупающих товары со
скидками "Small" и "Medium"



Взрослых клиентов,
покупающих **одежду**
премиального бренда

Перспективы

Изменить логику сбора данных датасета:



Policy Implication

Настройка **индивидуальных рекомендаций товаров** для каждого клиента в зависимости от его **типа устройства**

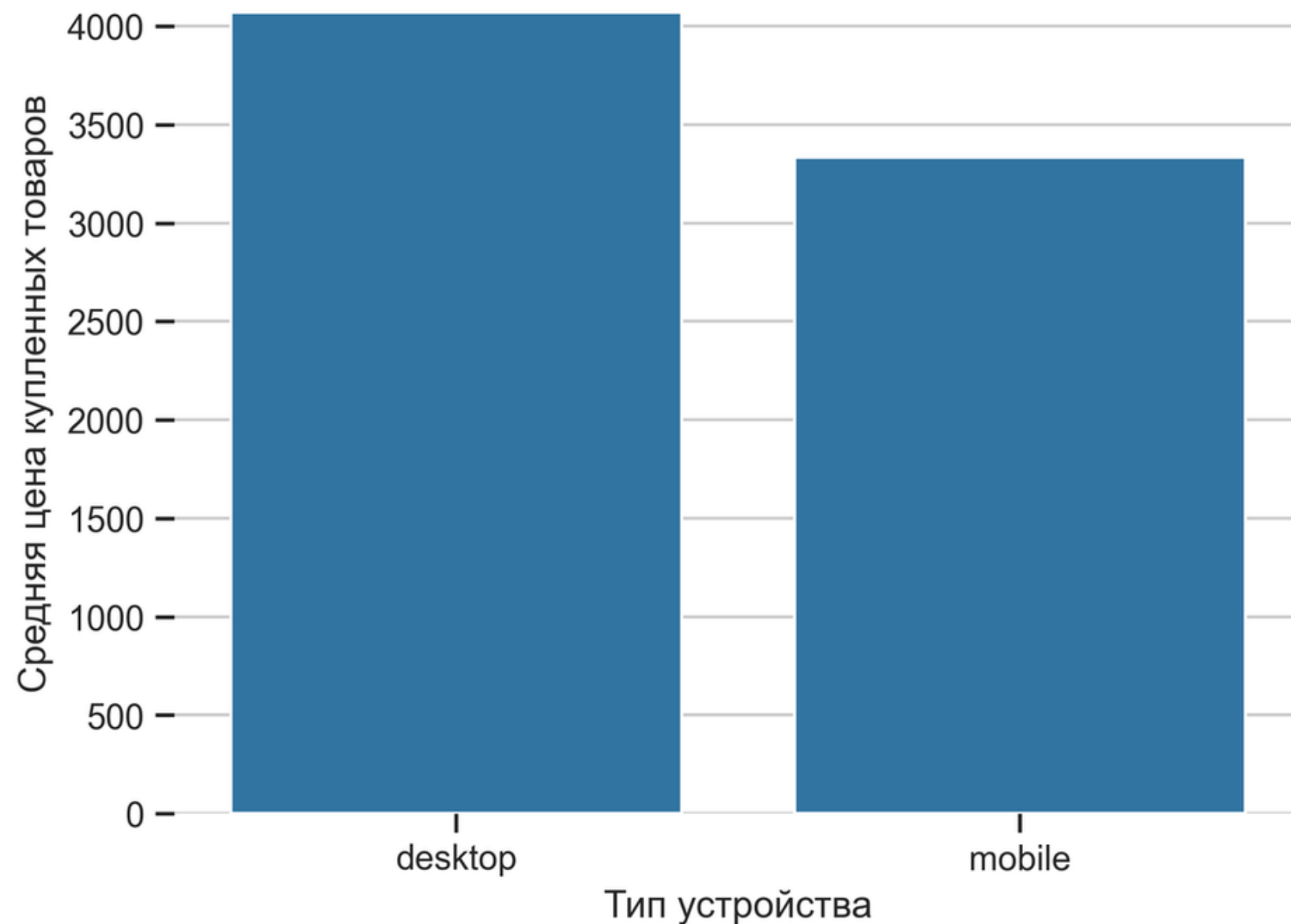


- Кратковременные персонализированные **скидки**
- Услуга **сравнения** товаров магазина для **удержания** пользователя на сайте
- Учёт особенностей платформы рекомендательными алгоритмами
- Услуга показа других товаров со **схожей** ценой и/или скидкой при просмотре **дорогих** товаров

Приложение

Математическое подтверждение механизма

Средняя цена купленных товаров по типам устройств



Используем t-критерий Уэлча

Цель: доказательство **различия** средней **цены** купленных товаров по типам устройств

Почему t-критерий Уэлча ?

Сравнение двух **несвязных** **выборок** с **различной** дисперсией

H0

Между средними групп **отсутствует** статически значимое отличие

H1

Средняя цена по типам устройств различается

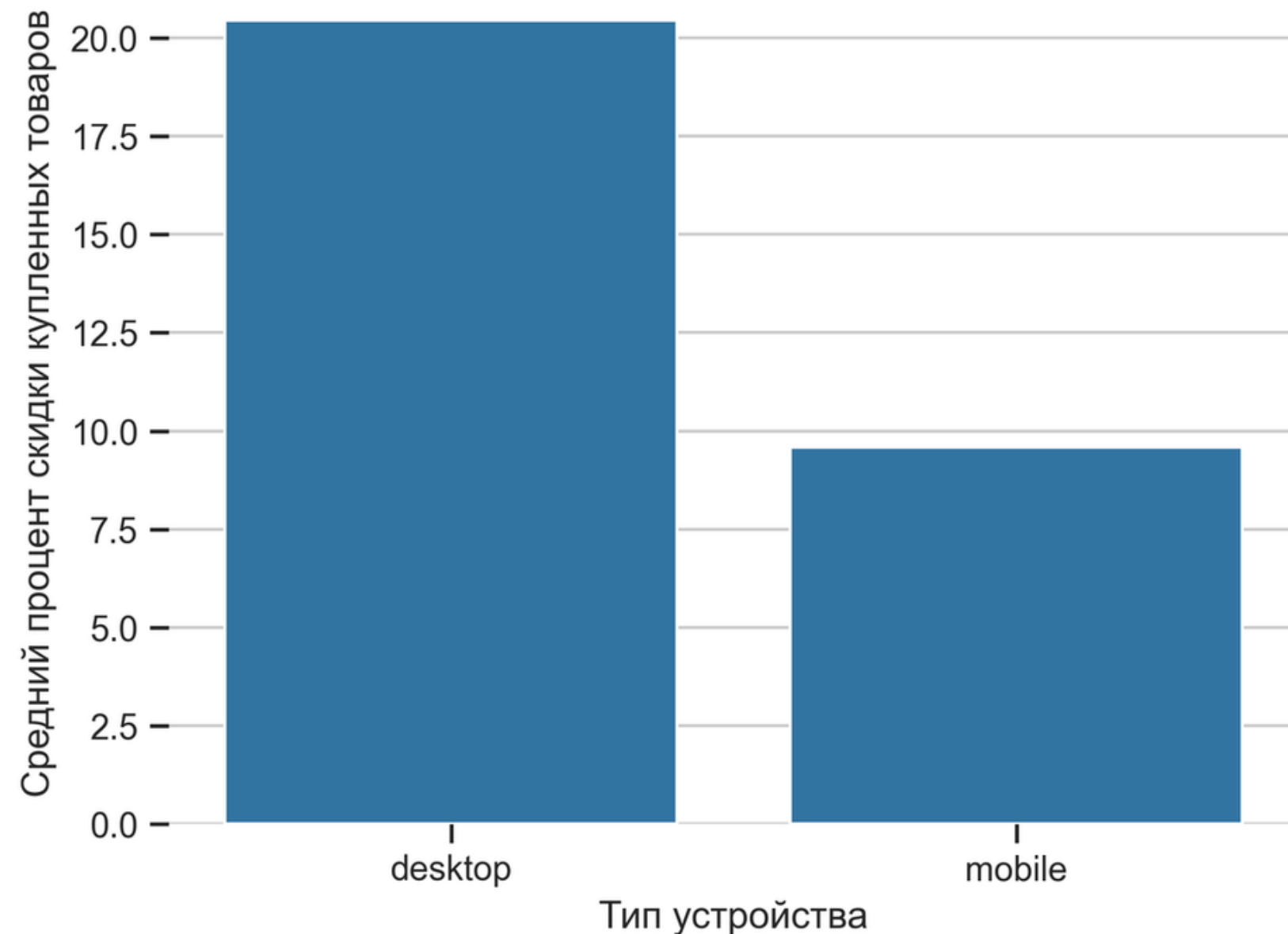
Результат: **p-value** ≈ 0

Вывод

- **p-value < 0.05**, нулевая гипотеза **отвергается**
- Средняя цена **различается** по типам устройств

Математическое подтверждение механизма

Средний процент скидки купленных товаров по типам устройств



Используем t-критерий Уэлча

Цель: доказательство **различия** среднего **процента скидки** купленных товаров по типам устройств

Почему t-критерий Уэлча ?

Сравнение двух **несвязных** выборок с **различной дисперсией**

H0

Между группами **отсутствует** статически
значимые отличия

H1

Средний процент скидки в группах типов устройств
различается

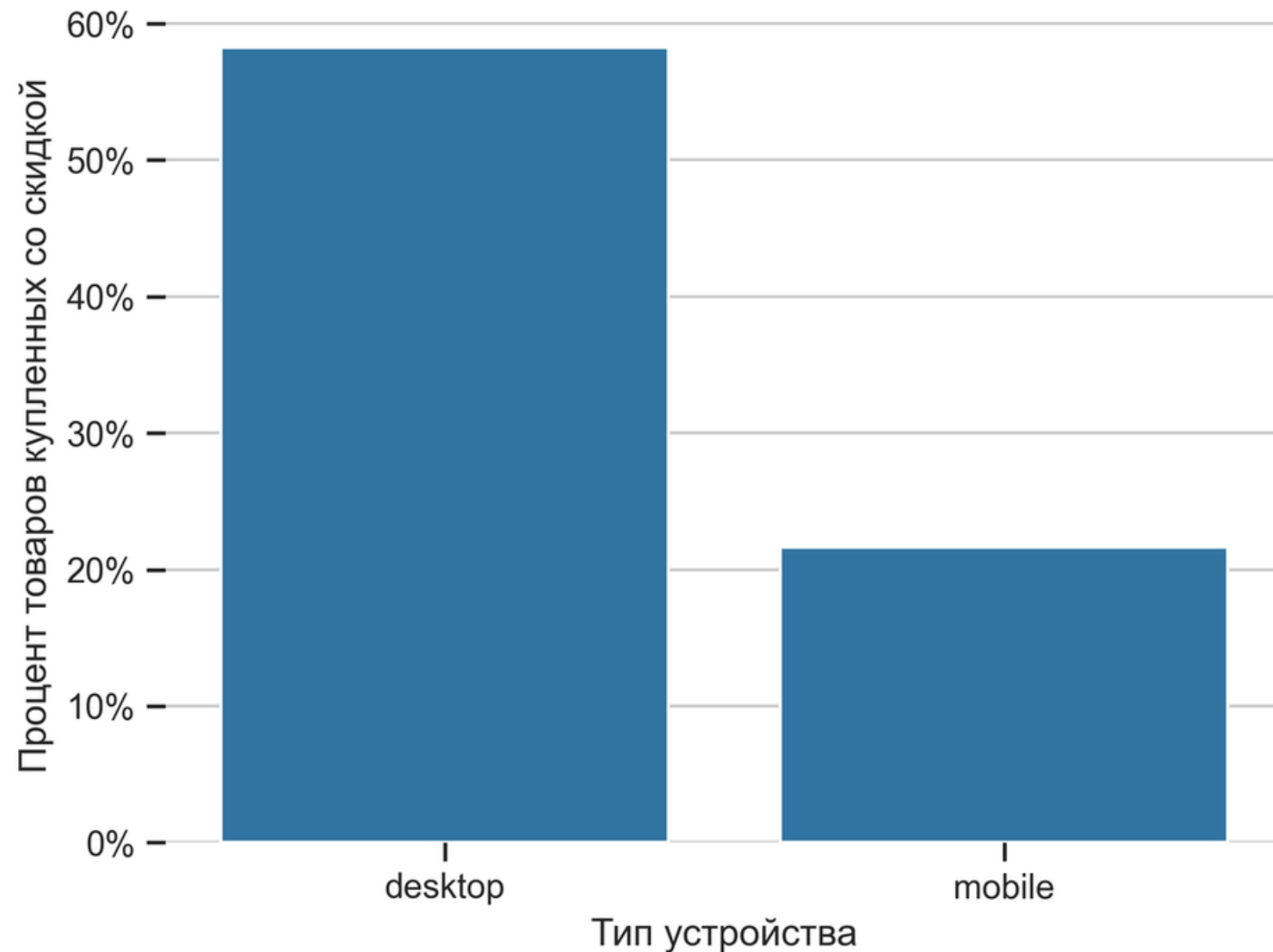
Результат: **p-value** ≈ 0

Вывод

- **p-value < 0.05**, нулевая гипотеза **отвергается**
- Средний процент скидки **различается** по типам устройств

Статистическое подтверждение механизма

Процент товаров купленных со скидкой по типам устройств



Используем **Хи-квадрат Пирсона** для доказательства различия доли товаров купленных **со скидкой** по **типам устройств**

H0

В генеральной совокупности **нет статистических различий процентов товаров купленных со скидкой** по типам устройств

H1

В генеральной совокупности **есть статистическое различие** процентов товаров купленных со скидкой по типам устройств

Статистическое подтверждение

	Покупка	Покупка со скидкой
Desktop	19 637	11 434
Mobile	314 210	67 968

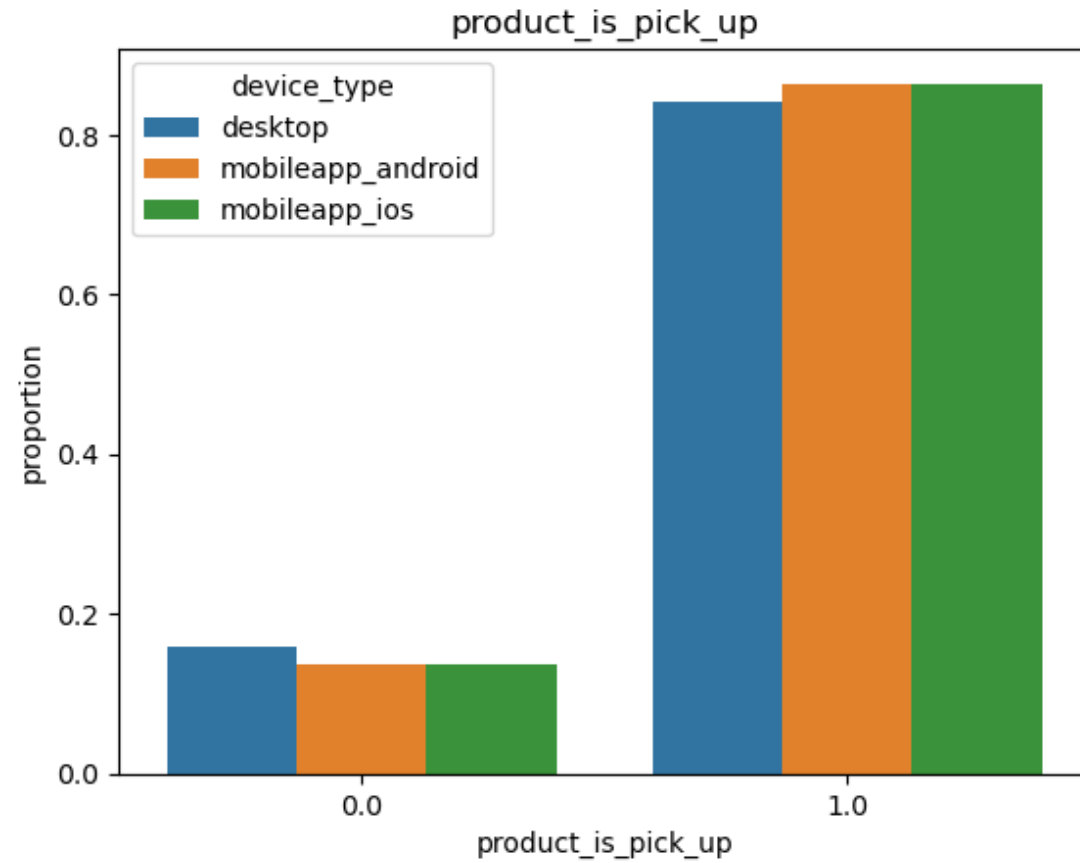
Результат:

chi2 6692

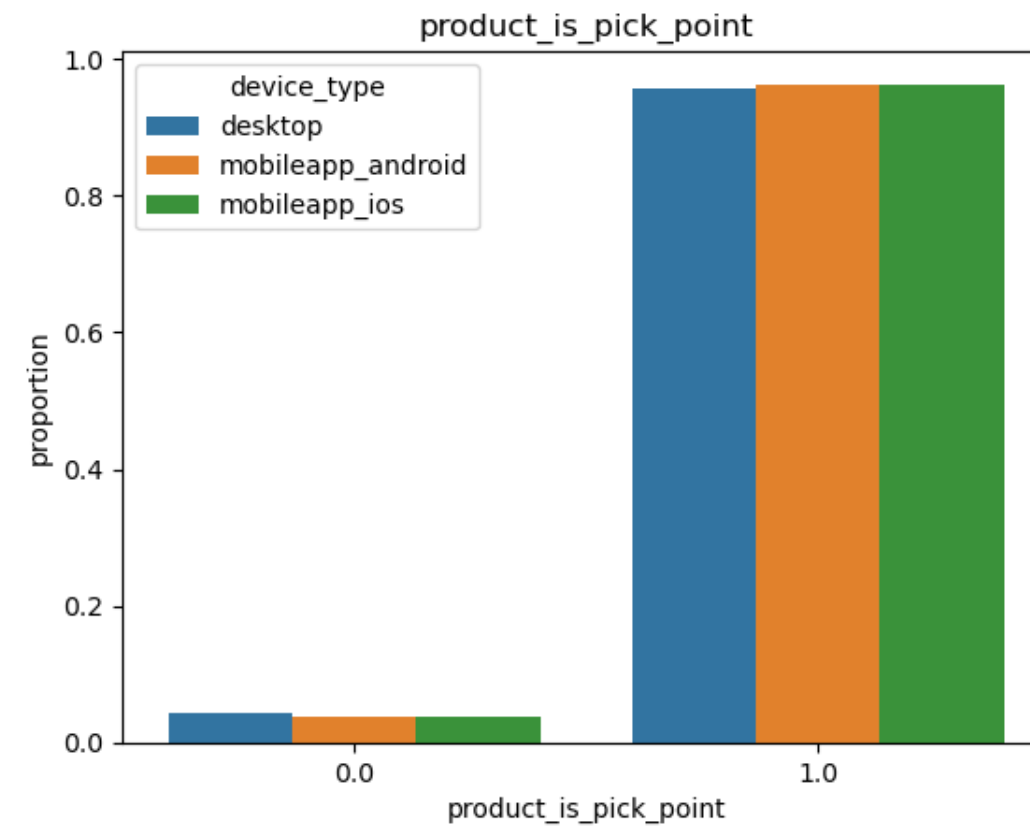
p-value 0

$p\text{-value} < 0.05$, **нулевая гипотеза отвергается**, процент товаров купленных со скидкой зависит от типа устройства

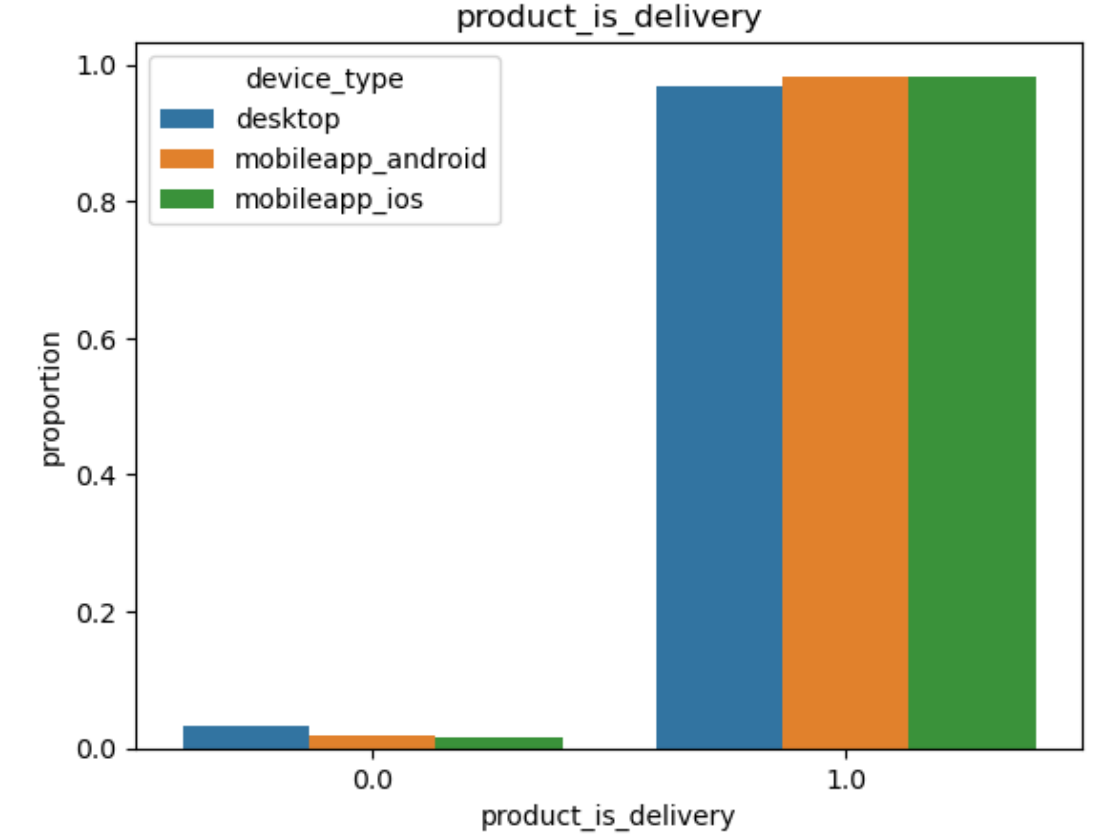
07 Приложение Пропорции достаточно больших выборок (от 15 000 записей)



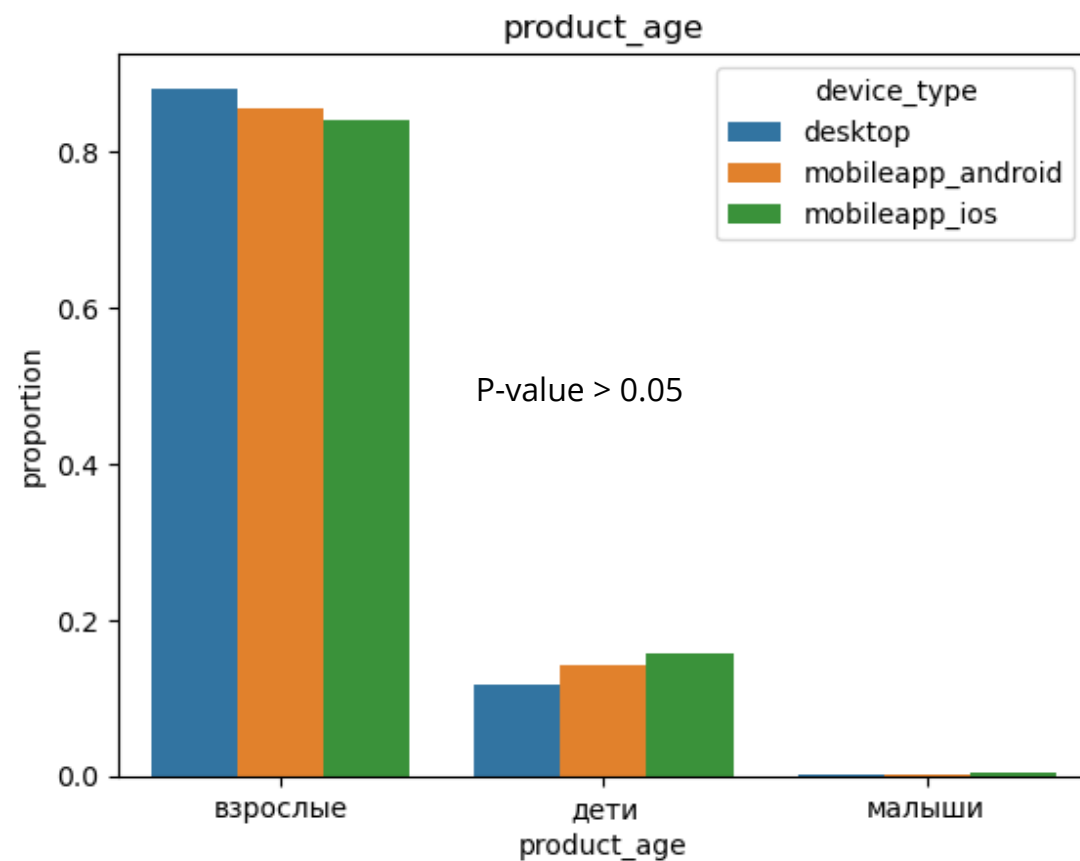
P-value > 0.05



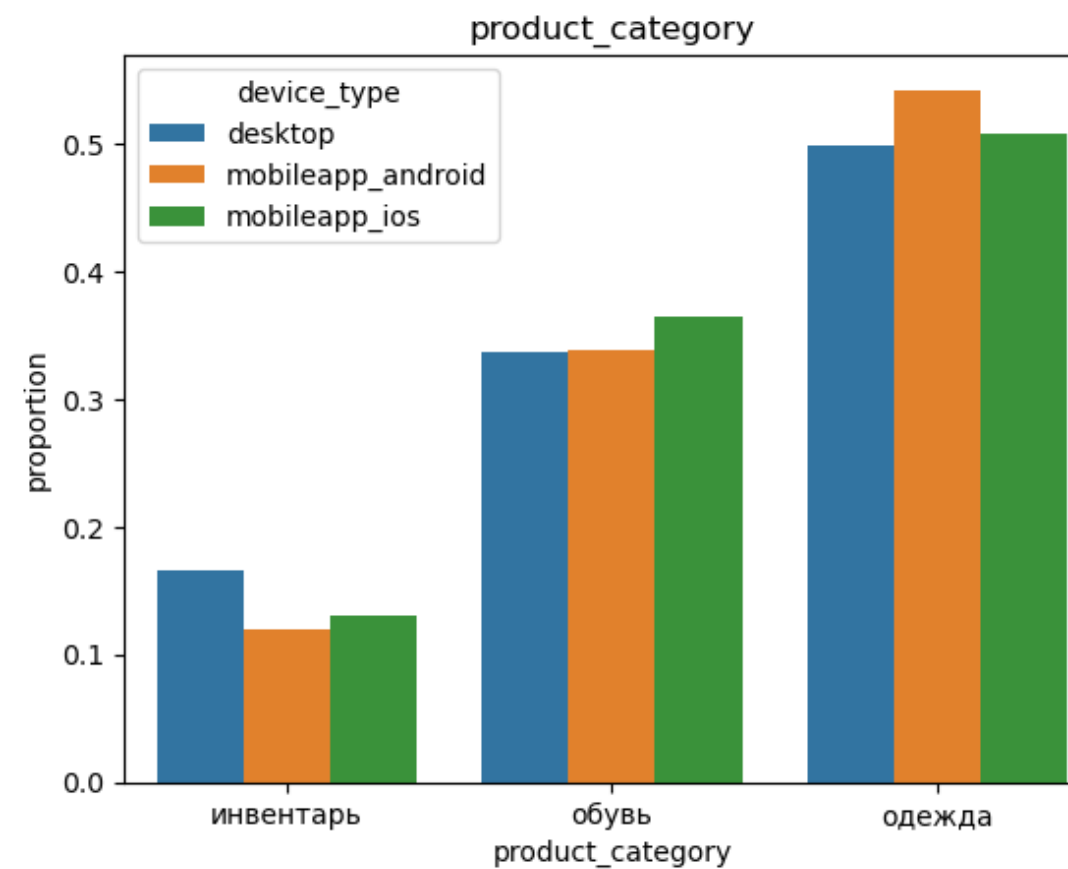
P-value > 0.05



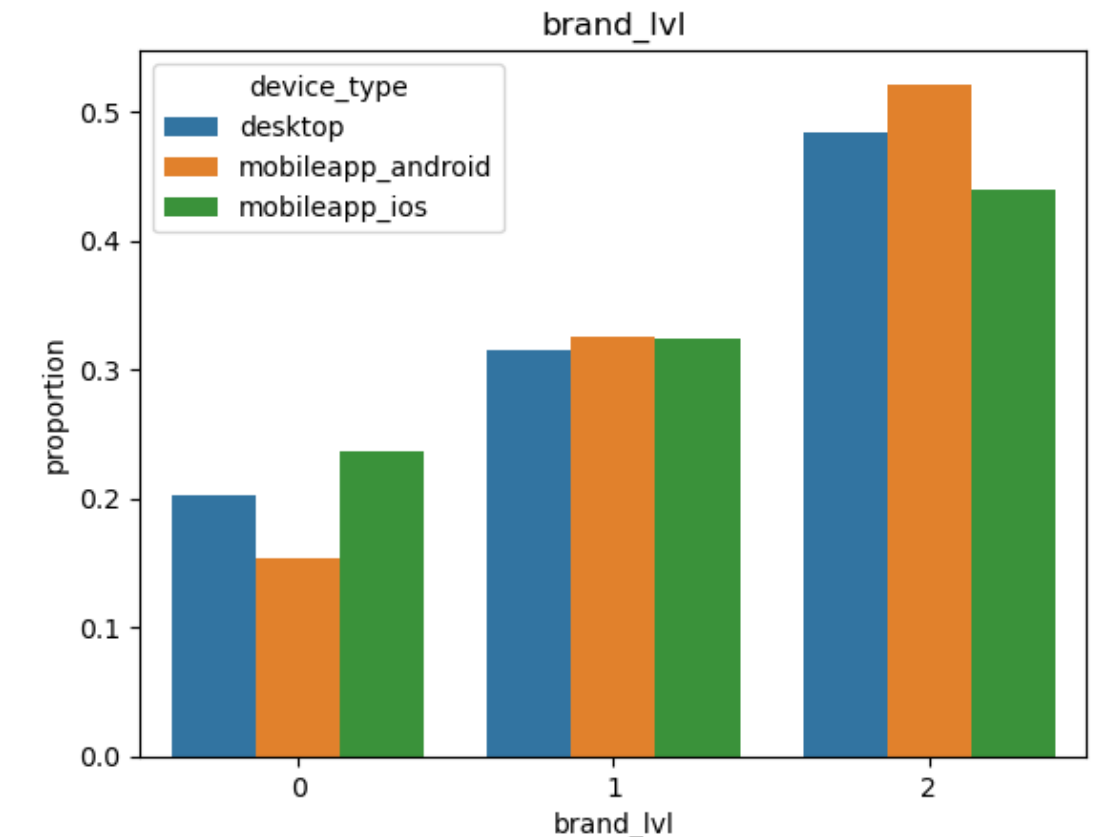
P-value > 0.05



P-value > 0.05



P-value > 0.05



P-value < 0.05. Различие обусловлено разницей в



В генеральной совокупности **нет**
статистических **различий** конверсий
по категориям цены



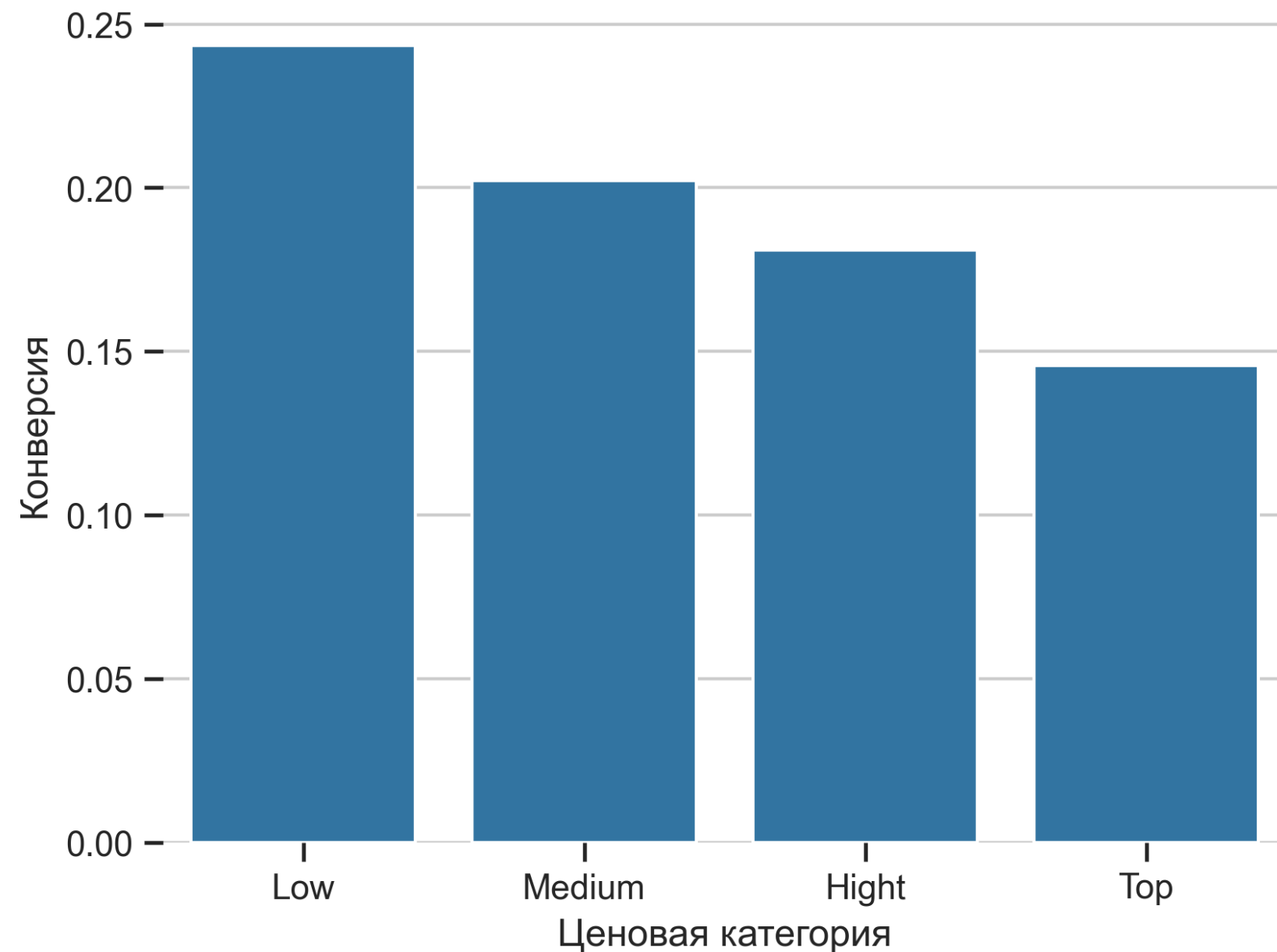
В генеральной совокупности **есть**
статистическое **различие** конверсий
по категориям цены

5%

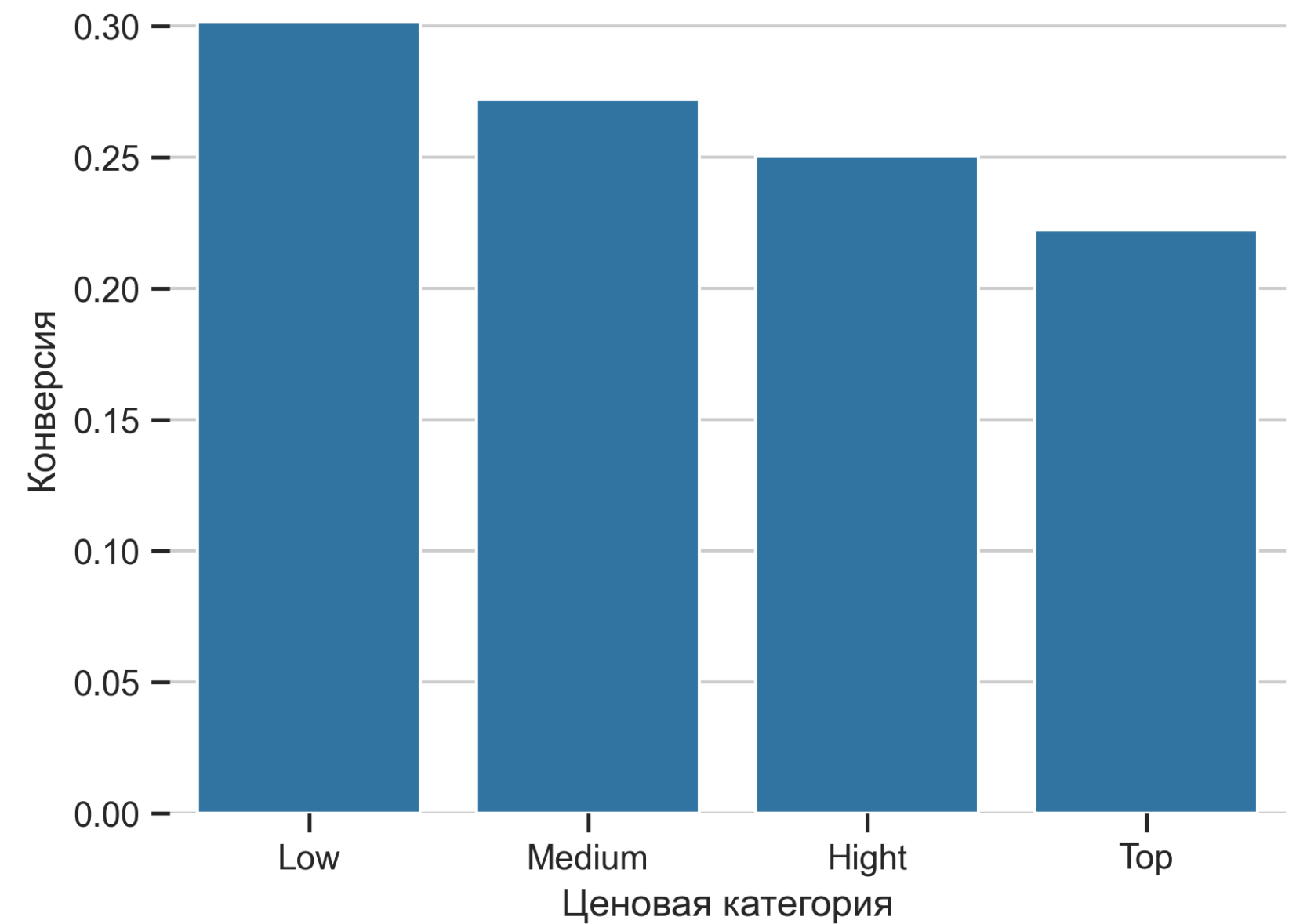
Уровень доверия

Используем Хи-квадрат Пирсона для доказательства влияния категории цены на конверсию на обоих этапах воронки

Из просмотра в добавление в корзину



Из корзины в покупку



10 Хи-квадрат Пирсона для доказательства влияния категории цены на конверсию на обоих этапах воронки

	Просмотр	Добавление в корзину
Low	1 620 975	394 507
Middle	1 672 315	338 134
Hight	1 674 429	302 842
Top	1 455 898	211 983

	Добавление в корзину	Покупка
Low	394 507	118 955
Middle	338 134	91 926
Hight	302 842	75 831
Top	211 983	47 135

chi2 = 33446

p-value ≈ 0

chi2 = 2916

p-value ≈ 0

р-value < 0.05 в обоих случаях, **нулевая гипотеза отвергается**, конверсии на обоих этапах зависят от типа устройства

Хи-квадрат Пирсона для доказательства влияния типа устройства на конверсию в разрезе ценовых категорий

Просмотр – Добавление в корзину

Price category	chi2	p-value
Low	3 397	< 0.05
Mid	2 553	< 0.05
Hight	2 910	< 0.05
Top	2867	< 0.05

Добавление в корзину - Покупка

Price category	chi2	p-value
Low	86	< 0.05
Mid	140	< 0.05
Hight	250	< 0.05
Top	73	< 0.05

р-value < 0.05 во всех случаях, **нулевая гипотеза отвергается**, конверсии в разрезе ценовых зависят от типа устройств



В генеральной совокупности **нет** статистических **различий** конверсий по категориям скидки



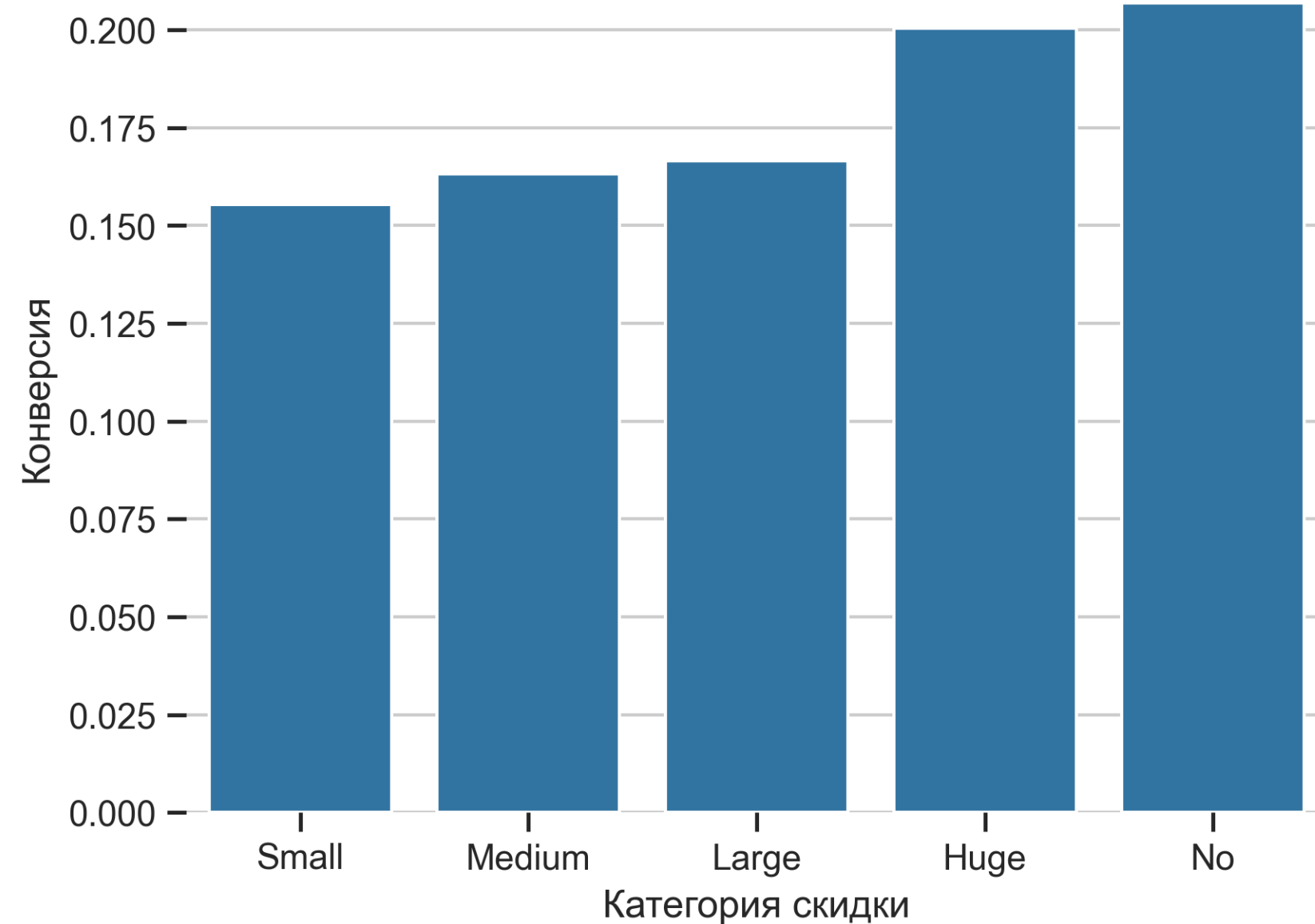
В генеральной совокупности **есть** статистическое **различие** конверсий по категориям скидки

5%

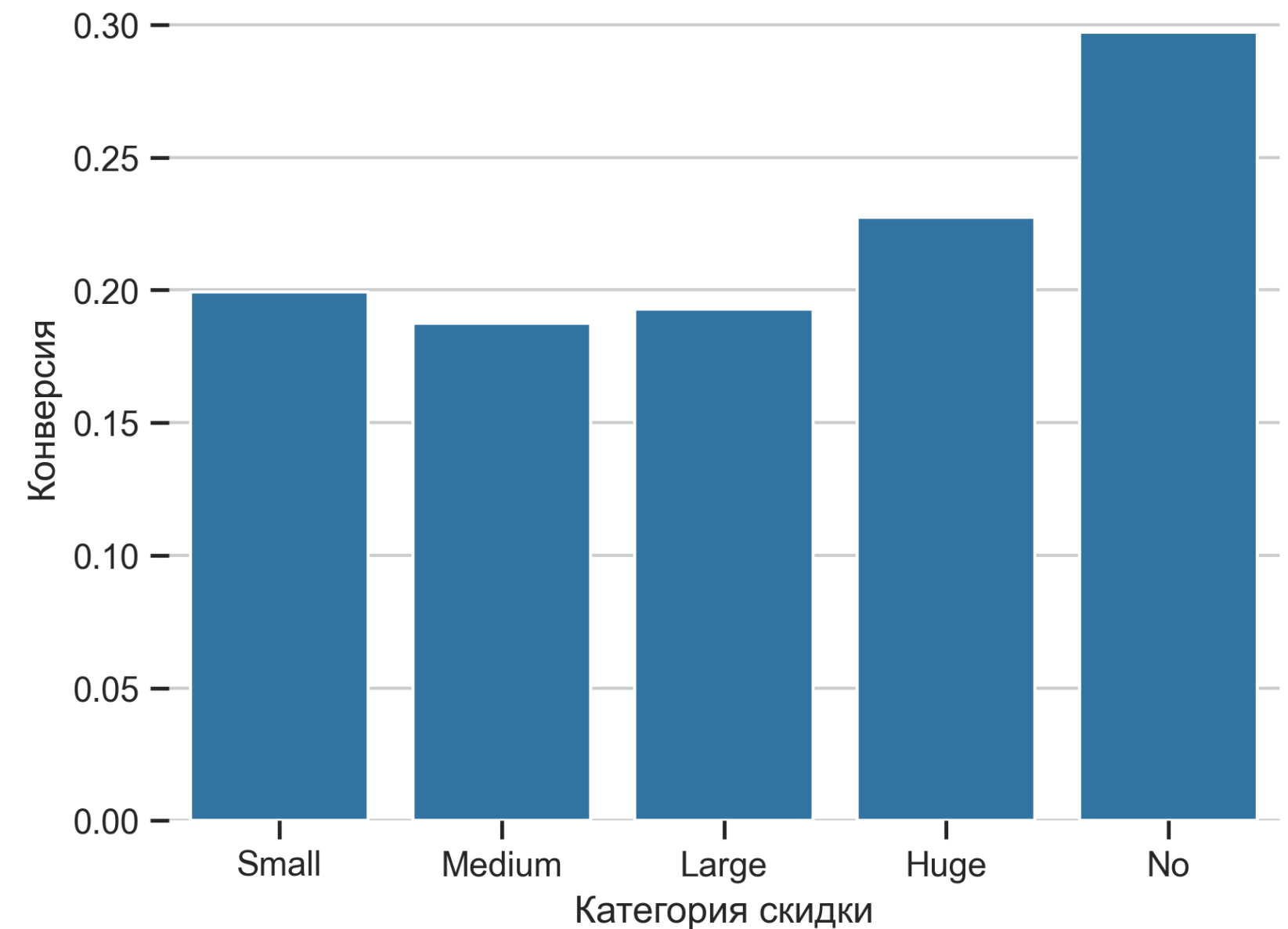
Уровень доверия

Используем Хи-квадрат Пирсона для доказательства влияния категории скидки на конверсию на обоих этапах воронки

Из просмотра в добавление в корзину



Из корзины в покупку



14 Хи-квадрат Пирсона для доказательства влияния категории скидки на конверсию на обоих этапах воронки

	Просмотр	Добавление в корзину
Small	573 392	89 074
Medium	600 445	97954
High	542 379	90 347
Top	567 543	113 791
No	4 139 858	856 300

	Добавление в корзину	Покупка
Small	89 074	17 745
Medium	97 954	18 352
High	90 347	17 422
Top	113 791	25 883
No	856 300	254 445

chi2

 = 11170

p-value

 ≈ 0

chi2

 = 7675

p-value

 ≈ 0

p-value < 0.05 в обоих случаях, **нулевая гипотеза отвергается**, конверсии на обоих этапах зависят от типа устройства

Хи-квадрат Пирсона для доказательства влияния типа устройства на конверсию в разрезе ценовых категорий

Просмотр – Добавление в корзину

Price category	chi2	p-value
Low	1508.87	< 0.05
Mid	863.92	< 0.05
High	934	< 0.05
Top	886.4	< 0.05
No	5578.7	< 0.05

Добавление в корзину - Покупка

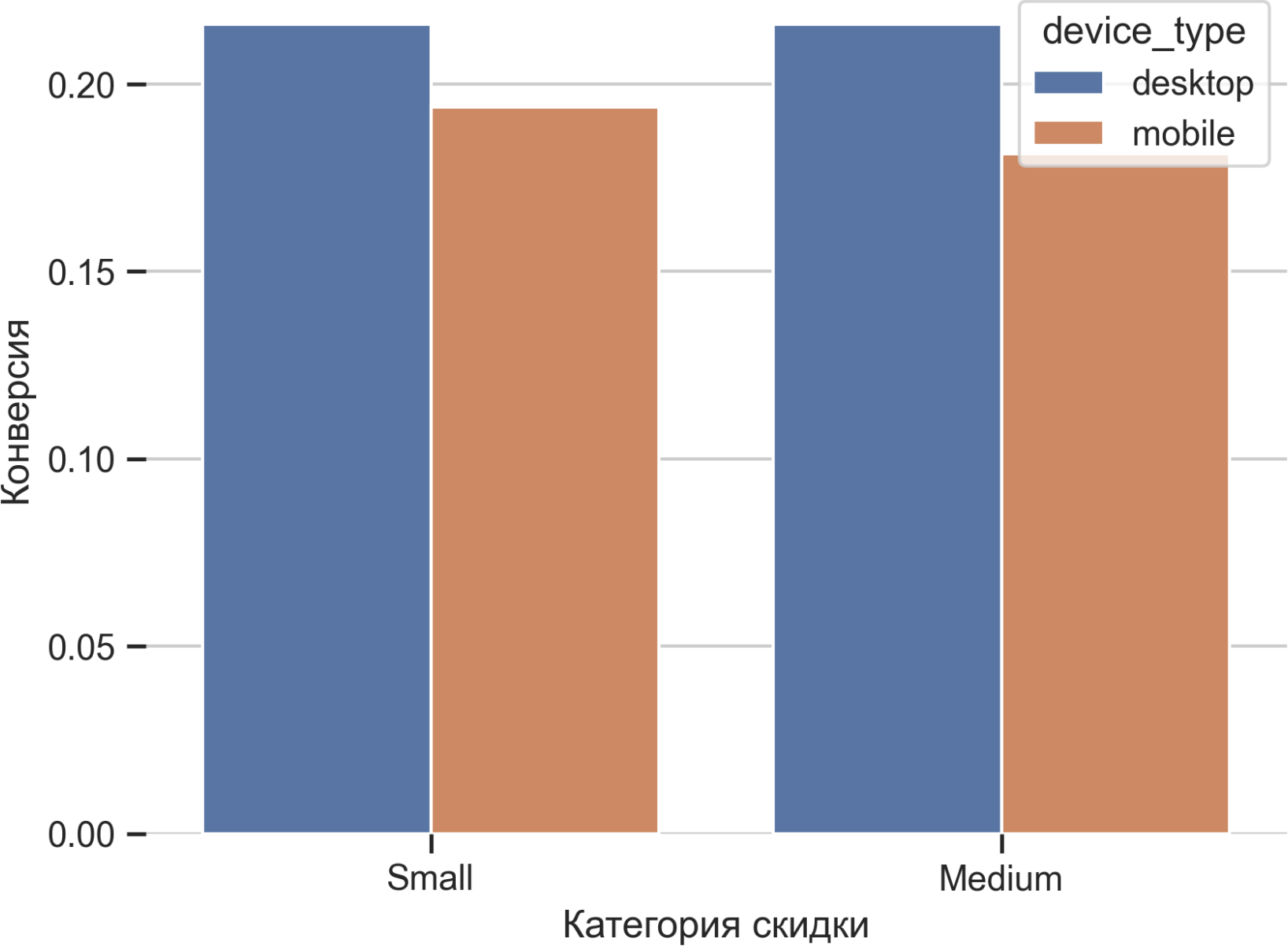
Price category	chi2	p-value
Low	33.601	< 0.05
Mid	72.66	< 0.05
High	4.17	< 0.05
Top	54.43	< 0.05
No	282.46	< 0.05

р-value < 0.05 во всех случаях, **нулевая гипотеза отвергается**, конверсии в разрезе ценовых зависят от типа устройств

Выборка категорий скидки 'Small' и 'Medium'

Конверсия из корзины в покупку по типу устройства

Используем **Хи-квадрат Пирсона** для доказательства различия конверсий по типам устройств на выборке скидок "Small" и "High"



Категория скидки Small

	Покупка	Добавление в корзину
Desktop	4 710	21 806
Mobile	13 035	67 268

chi2 = 33.6

p-value ≈ 0

Категория скидки Medium

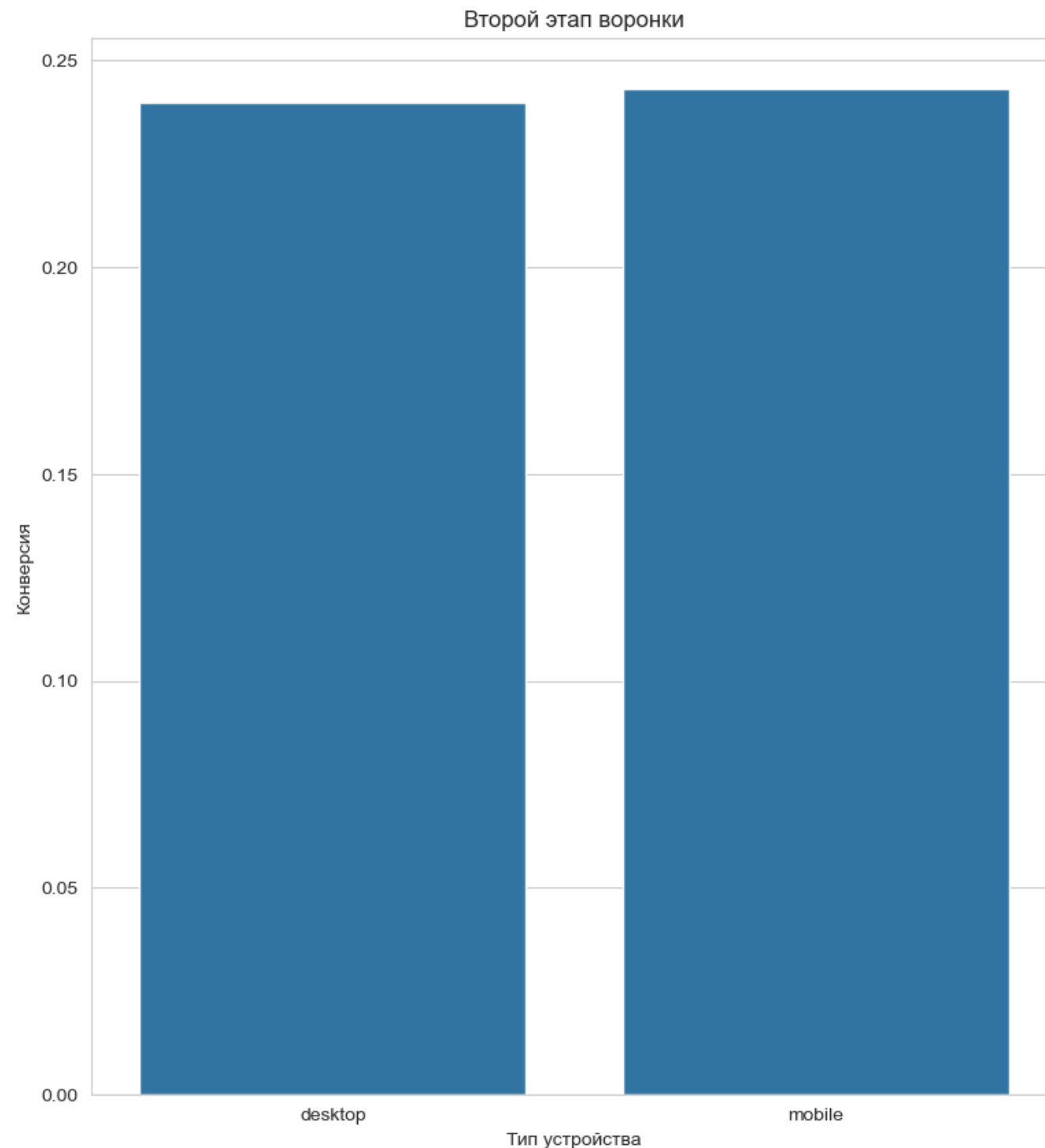
	Покупка	Добавление в корзину
Desktop	3 618	16 755
Mobile	14 734	81 199

chi2 = 72.7

p-value ≈ 0

Выборка категорий “Взрослая одежда премиальных брендов”

Конверсия из корзины в покупку
по типу устройства



H0

Между группами **отсутствуют** статически значимые отличия

H1

Конверсия второго этапа воронки **значимо различается** по разным типам устройства

Используем **Хи-квадрат Пирсона**

Результат:

chi2

0.199

p-value

0.655

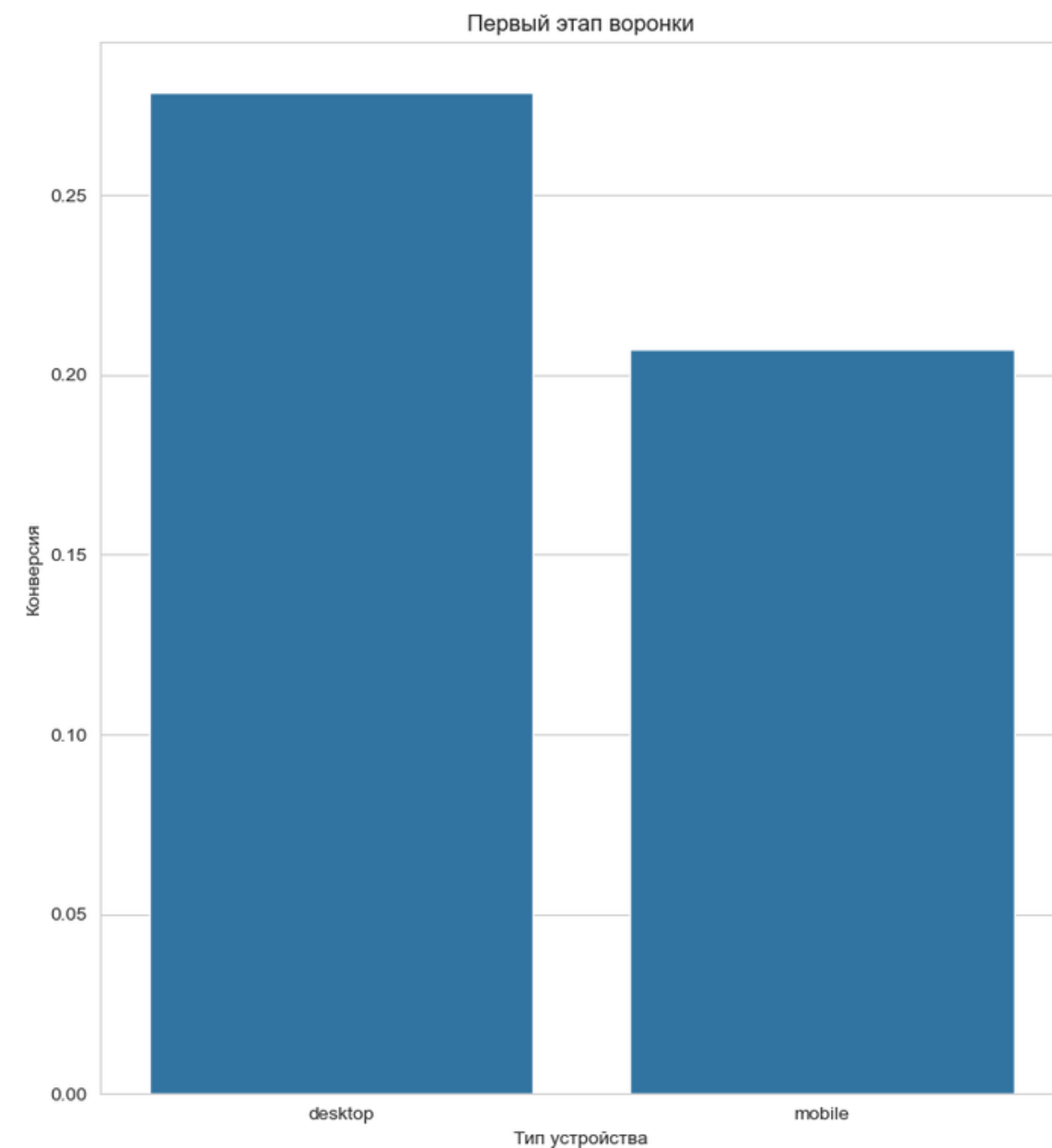
p-value > 0.05



В генеральной совокупности **нет** значимых отличий

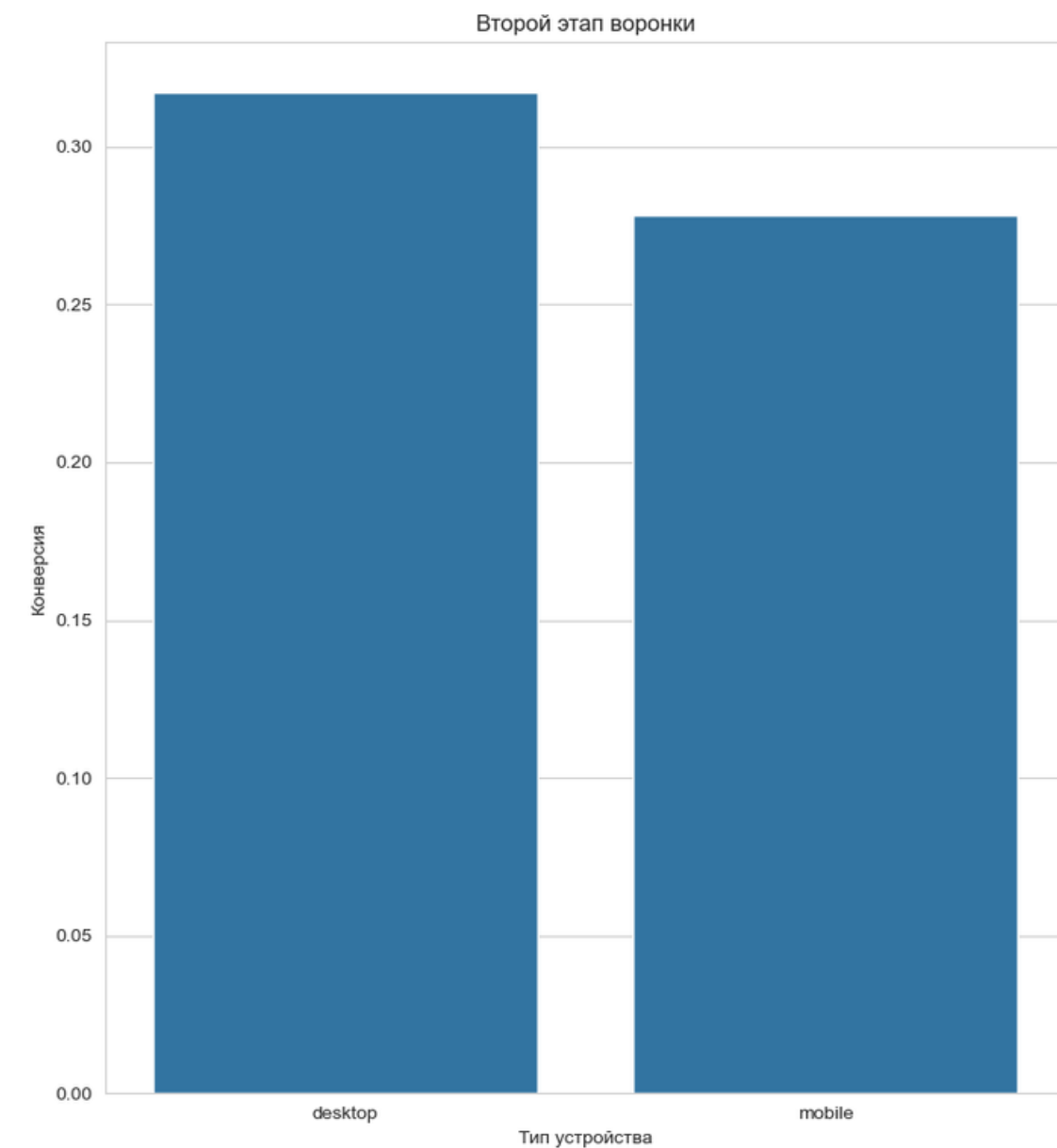
Выборка авторизованных пользователей

Конверсия из просмотра в корзину
по типу устройства



p-value < 0.05

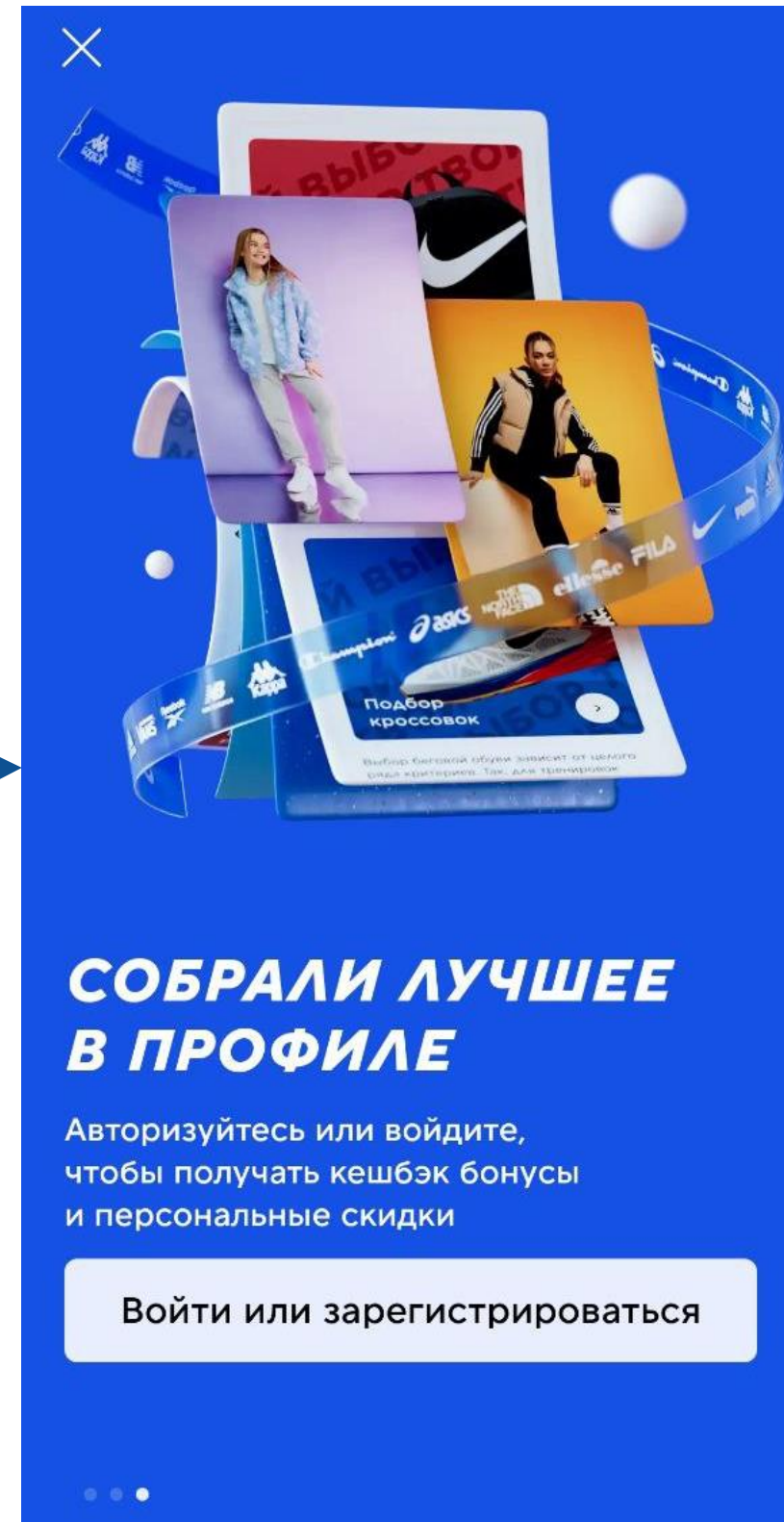
Конверсия из корзины в покупку
по типу устройства



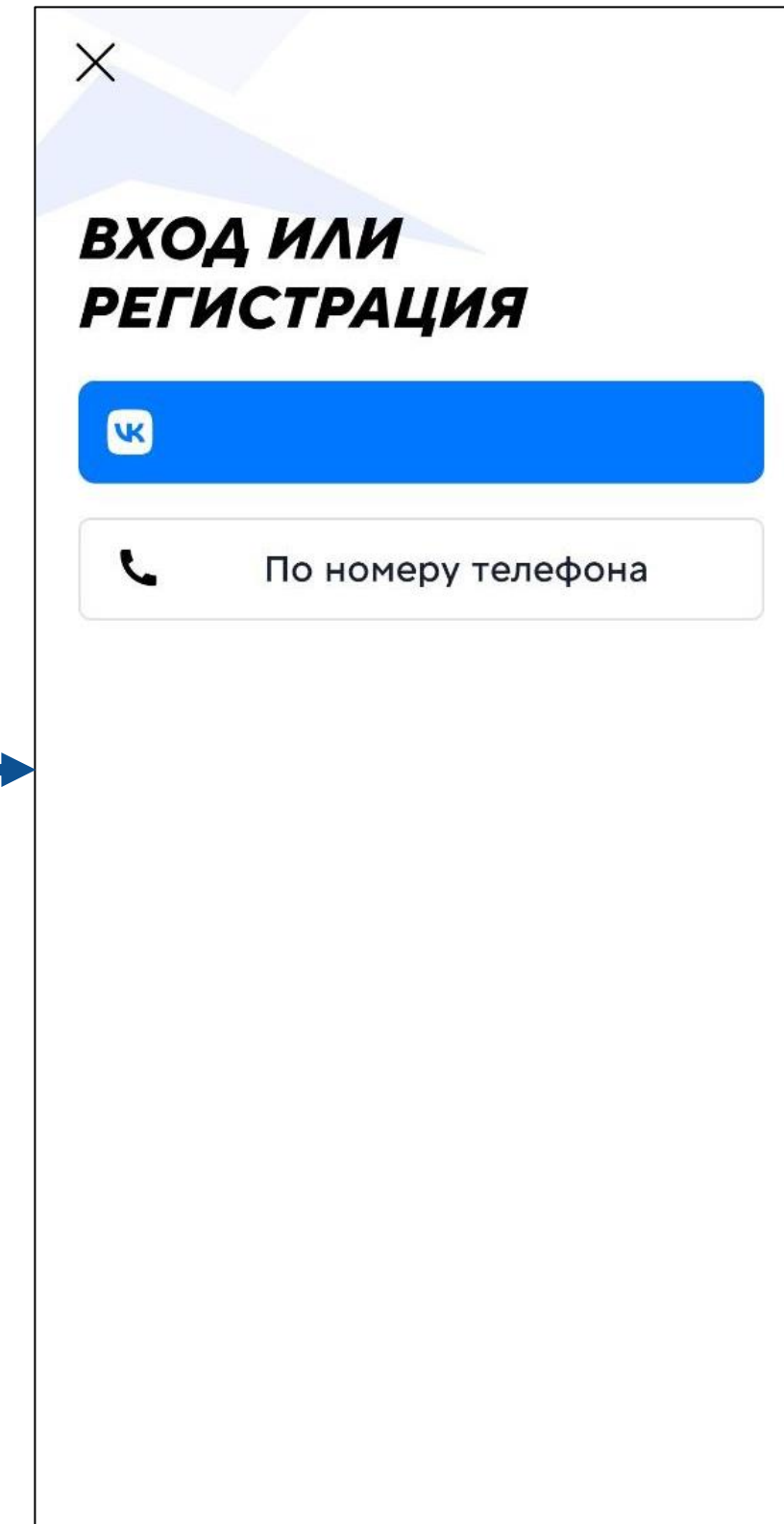
p-value < 0.05

Хи-квадрат Пирсона

При заходе в приложение



При закрытии
этого окна



Вылезает
только при
оформлении
заказа



×


**ВХОД ИЛИ
РЕГИСТРАЦИЯ**

Номер телефона*


+7


получить код


или

 Войти с VK ID

Что даёт авторизация?


Кэшбек бонусами с
каждой покупки


Персональные
цены и скидки


Быстрый заказ:
сохраняем адреса
получения

Продолжить без авторизации