

# Первичная верификация расчётов оценщиков

Ylisjärvi: временный инструмент автоматизированного  
контроля корректировочных таблиц оценщиков

Мурашев К. А.<sup>1</sup>

<sup>1</sup>ООО "Совконсалт"  
директор

30 мая 2024 г.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

# Содержание I

- 1 Используемые обозначения
- 2 История разработки
- 3 Постановка задачи
- 4 Общее описание метода корректировок
- 5 Основные проблемы расчётов
- 6 Теоретические требования к расчётам
  - Соотношение числа аналогов и корректировок
  - Первичный анализ данных локального конкурентного сегмента
  - Типы данных в оценке

## Содержание II

- Проверка гипотез о распределении непрерывных переменных
- Обработка данных, характеризующих отношение
- Обработка категориальных переменных
- Обработка бинарных данных
- Формирование итогового суждения о необходимости применения конкретных корректировок

### 7 Методы и метрики проверки готовых расчётов

- Визуальный анализ
- Проверка изменения распределения цен аналогов

## Содержание III

- Проверка устранения связи между ценой и переменными
- 8 Методы валидации и тестирования
  - Валидация
  - Итоговое тестирование
- 9 Практическая реализация на уровне MVP
  - Пройденный путь
  - Печальный финал
  - Новые возможности

# Используемые обозначения I

- [...] — Округление до ближайшего целого:  $8.8 \rightarrow 9$ .
- [...] — Округление до ближайшего меньшего целого:  
 $8.8 \rightarrow 8$ .
- [...] — Округление до ближайшего большего целого:  
 $8.8 \rightarrow 9$ .

# История разработки I

- Первоначальная задача была поставлена одним из российских банков: выявление запредельных оценок без глубокого погружения в отчёт
- Большая часть оценок выполнялась с помощью метода внешних корректировок, а качество оценки вызывало недовольство

## История разработки II

- Реализация заключалась в разработке формально-логических количественных критериев проверки расчётов оценщиков, а также программного инструмента для выполнения этой проверки
- Поскольку не все предполагаемые пользователи хорошо владели методами оценки и анализа данных, задача стояла в разработке максимально простых, местами примитивных, методов проверки.

## История разработки III

- Заранее прошу прощение за некоторую примитивность методов.
- Основная область применения: оценка недвижимости
- Был разработан набор правил и back-end для анализа расчётной таблицы



## История разработки IV

- В период разработки необходимость в её завершении отпала, поскольку заказчик принял решение перейти на скоринговые модели оценки стоимости на основе машинного обучения и заказать уже их разработку, ещё более упростив задачу верификации отчётов внешних оценщиков
- Права на все разработки остались у разработчика, поэтому могут быть переданы в общественное использование

Используемые обозначения

История разработки

Постановка задачи

Общее описание метода корректировок

Основные проблемы расчётов

Теоретические требования к расчётам

Методы и метрики проверки готовых расчётов

Методы валидации и тестирования

Практическая реализация на уровне MVP

# История разработки V

- На данный момент готов основной набор правил, а также back-end на уровне MVP (Minimum Viable Product)

# Постановка задачи I

- Следует признать, что по состоянию на II квартал 2024 года метод внешних корректировок <sup>1</sup>
- Данный метод не является лучшим, он устарел, поскольку возник в эпоху деревянных счёт, арифмометров и объявлений в бумажных газетах (Джордж Делл: "по-прежнему используются методы и устаревшие решения, разработанные для скудных и труднодоступных данных 1930-х годов" — <https://georgedell.com/what-is-a-valuation-model>), и со временем будет заменён другими

## Постановка задачи II

- Тем не менее, на данном этапе именно этот метод является наиболее распространённым и поэтому к нему необходимо предъявлять требования во избежание запредельных либо просто некачественных оценок независимых оценщиков
- Поэтому необходима разработка ряда теоретических требований к методу
- Также необходимы конкретные количественные методы оценки качества расчёта во избежание споров и долгих разговоров

## Постановка задачи III

- Поскольку проверка чужой работы всегда достаточно трудозатратна, целесообразно её автоматизировать
- Следует отметить, что данная проверка представляет собой специализированный инструмент проверки конкретного расчёта оценщика, выполненного методом корректировок. Это не проверка самих корректировок, их источников и т. д. Оценщик сам выбирает, какой метод и какие источники данных и информации использовать

## Постановка задачи IV

- Данная проверка не включает в себя проверку обоснования выбора метода расчёта, выбора конкретных аналогов, сами корректировки, их обоснованность и источники
- Речь идёт только о проверке самого расчёта.

# Постановка задачи V

- Задача заключалась не только в контроле расчётов, но и применению правильной методологии проверки залоговой службой. Распространённой проблемой было непонимание, включение каких корректировок необходимо, а каких нет. Часто залоговики требовали включать корректировки просто по факту их существования в источнике без обоснования необходимости в конкретном случае.

---

<sup>1</sup>внешние корректировки — те, которые оценщик не получает в ходе собственного исследования, а берёт в готовом виде из внешних источников как то: публикации в ИТС Интернет, статьи, экспертные опросы, справочники и т. д.

# Общее описание метода корректировок I

## Общее описание метода корректировок

- Задача внесения корректировок есть минимизация отличия аналогов от объекта оценки
- Часть корректировок (например коэффициент перехода от цен оферт к ценам транзакций) объективно необходима, часть требуется только в случае необходимости в конкретном случае



# Общее описание метода корректировок II

## Общее описание метода корректировок

- Важно помнить, что используемые аналоги это не "какие-то" объекты, найденные в Интернете, это объекты, ценообразование которых соответствует ценообразованию в конкретном конкурентом сегменте рынка, к которому относится объект оценки
- Идеальный набор аналогов вообще не требует внесение каких-либо корректировок кроме коэффициента перехода от оферты к сделке

Используемые обозначения

История разработки

Постановка задачи

Общее описание метода корректировок

Основные проблемы расчётов

Теоретические требования к расчётам

Методы и метрики проверки готовых расчётов

Методы валидации и тестирования

Практическая реализация на уровне MVP

# Общее описание метода корректировок III

## Общее описание метода корректировок

- Задача оценщика подобрать такие аналоги, которые требуют минимальное внесение корректировок
- Внесение каждой корректировки требует обоснование (нулевая гипотеза всегда говорит о том, что различие отсутствует)
- Никто не обязан доказывать и обосновывать отсутствие той или иной корректировки

# Общее описание метода корректировок IV

## Общее описание метода корректировок

- Метод корректировок по умолчанию предполагает, что аналоги достаточно близки к объекту оценки, поэтому необходимость внесения корректировки требует обоснование, но не наоборот.
- Следует помнить главное: набор аналогов и корректировок это не просто набор строк и в таблице, это математическая модель ценообразования на конкретном локальном конкурентном сегменте рынка, хотя и несколько своеобразная.

# Общее описание метода корректировок V

## Общее описание метода корректировок

- Корректировка не является нормативной величиной, её нельзя воспринимать как безусловное руководство к действию. Это попытка количественного измерения восприятия участниками рынка тех или иных свойств объектов с точки зрения их влияния на стоимость.
- Анализ рынка, выявление локального конкурентного сегмента, к которому относится объект оценки во всех случаях является обязанностью оценщика.

# Общее описание метода корректировок VI

## Общее описание метода корректировок

- Поскольку корректировки рассчитываются в среднем по рынкам регионов или их групп либо по их крупным сегментами, а сами значения отражают изменение стоимости некоего усреднённого объекта, необходимость их внесения не очевидна априорно на основании одного только факта, что у оценщика есть значения признаков аналогов, а в справочнике есть корректировка для этого признака.

Используемые обозначения

История разработки

Постановка задачи

Общее описание метода корректировок

Основные проблемы расчётов

Теоретические требования к расчётам

Методы и метрики проверки готовых расчётов

Методы валидации и тестирования

Практическая реализация на уровне MVP

# Общее описание метода корректировок VII

## Общее описание метода корректировок

- Внесение любой корректировки может быть проведено только при одновременном наличии двух оснований для этого:
  - профессиональное мнение оценщика и его опыт указывают на невозможность учёта различия в конкретном признаке на конкретном рынке;
  - статистические тесты, выполненные на конкретном наборе аналогов, указывают на необходимость отклонения гипотезы об отсутствии влияния признака на стоимость

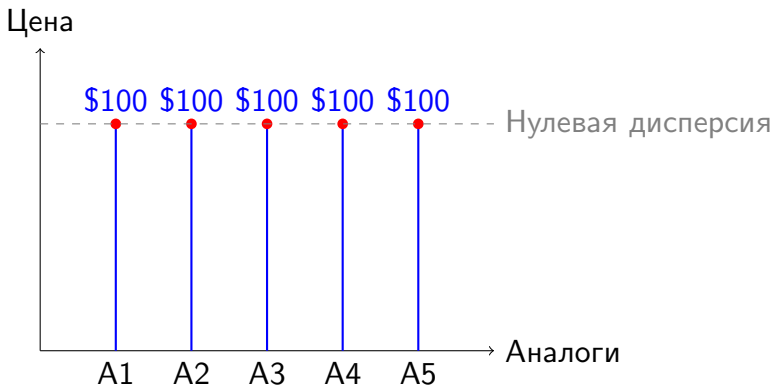
# Общее описание метода корректировок VIII

## Общее описание метода корректировок

- Итогом внесения корректировок должно быть явное уменьшение отличий между аналогами, а также между ними и объектом оценки.
- Следует контролировать направление и количественные меры корректировок. В идеале, корректировки нормально распределены со средним равным 0. Данное требование не всегда достижимо.
- Тем не менее, контроль за направлением корректировок необходим во всех случаях.

- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

# Идеальные аналоги I

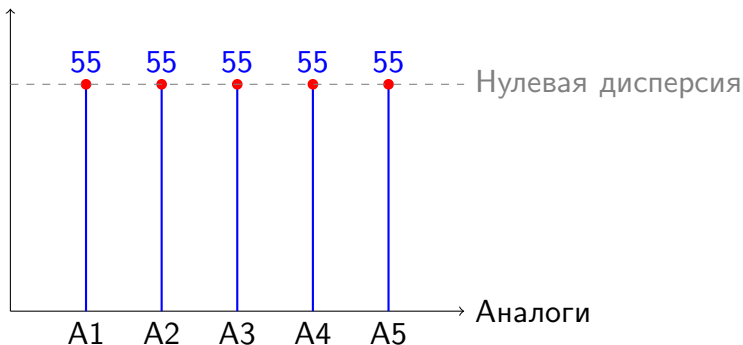




- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

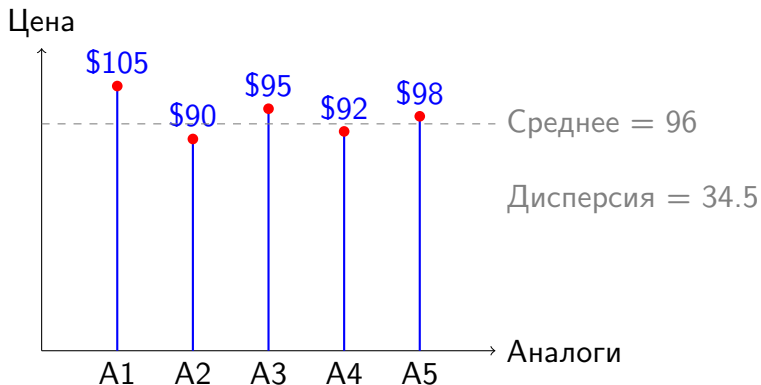
## Идеальные аналоги II

Характеристика



- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

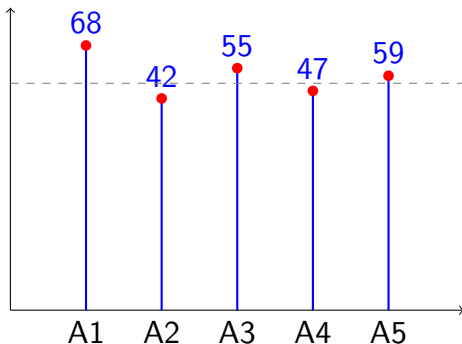
# Реальные аналоги I



- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

## Реальные аналоги II

Характеристика



Среднее = 54

Дисперсия = 103.7

Объект оценки = 61

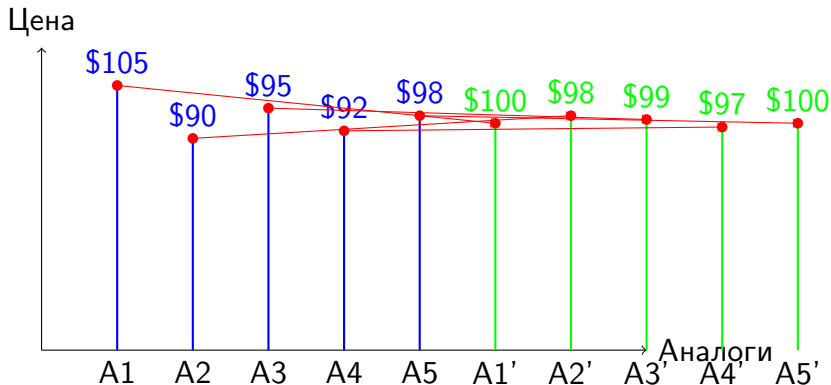
Аналоги

# Процесс корректировок I

- В результате применения корректировок оценщик снижает цену аналогов, имеющих значение характеристики "лучше", чем у объекта оценки, и повышает тех, у кого оно "хуже".
- В результате применения корректировок снижается дисперсия цены, в рассматриваемом примере оно снизилось с 34.5 до 1.7, среднее при этом выросло с 96 до 99, что отражает факт того, что объект оценки "лучше" 4/5 аналогов.

- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

## Процесс корректировок II



# Основные проблемы расчётов I

- Недостаточное количество аналогов. Следует ещё раз сказать, что аналоги это не просто "какие-то объекты". Набор аналогов есть проекция конкурентного сегмента рынка в удобную для работы форму. Их количество должно быть разумно достаточным для моделирования ценообразования.
- Конкретное количество аналогов зависит от рынка и практических возможностей поиска данных на нём, а также количества используемых корректировок.

## Основные проблемы расчётов II

- Можно сказать иначе: количество допустимых к применению корректировок зависит от располагаемого числа аналогов, таким образом часто имеет место проблема слишком большого числа применяемых корректировок.
- Выбор в качестве аналогов объектов, имеющих заведомо неразумную меру отличия относительно объекта оценки и между собой (например, отличие площади в 30 раз).

## Основные проблемы расчётов III

- Внесение конкретной корректировки без проверки необходимости её применения: внесение корректировки это вынужденная мера тогда, когда уже невозможно игнорировать отличие между аналогом и объектом оценки.
- Желание внести корректировку "просто так", на основании того, что она присутствует в источнике, есть ошибка.



## Основные проблемы расчётов IV

- Внесение корректировок таким образом, что итоговая мера различия между аналогами и объектом оценки, а также между аналогами, не уменьшается до приемлемого уровня.
- Внесение дублирующих корректировок, имеющих одно и то же фундаментальное содержание (если есть корректировка на отдельный вход, то корректировка на первый этаж скорее всего не нужна, если только не доказано, что её расчёт проводился строго для помещений на первом этаже).

## Основные проблемы расчётов V

- Внесение корректировок таким образом, что общее абсолютное значение корректировки превышает исходное значение стоимости аналога.
- Внесение корректировок таким образом, что общее абсолютное значение суммы корректировок, если бы они применялись по отдельности, превышает исходное значение стоимости аналога.
- Внесение более 2 корректировок к одному аналогу.

## Основные проблемы расчётов VI

- Общей проблемой любого расчёта, выполненного на конкретной выборке аналогов, независимо от метода (это могут быть корректировки, индивидуальный расчёт с помощью регрессионной модели, случайный лес, XGBoost, нейронная сеть т. д.), является то, что модель хорошо работает только на той выборке аналогов, на которой её строили.

## Основные проблемы расчётов VII

- При этом нас интересует ценообразование не конкретной выборки (мы и так знаем о ней всё), а всего локального сегмента, к которому относится объект оценки. Для проверки поведения расчётной модели, в т.ч. корректировок, необходимо её тестирование на части найденных аналогов, не включённых в расчётную модель.

## Основные проблемы расчётов VIII

- В этом случае, мы может сделать осторожные предположения о применимости модели для её использования при решении практических задач оценки стоимости.

## Соотношение числа аналогов и корректировок I

- Число аналогов ( $n$ ) и корректировок ( $k$ )<sup>2</sup> должно соответствовать друг другу. Минимальное число аналогов (допускается при работе на малоактивных рынках):

$$n = \left\lceil \frac{k + 3}{0.8} \right\rceil \quad (1)$$

## Соотношение числа аналогов и корректировок II

Рекомендуемое число аналогов (используется в остальных случаях):

$$n = \left\lceil \frac{2k + 3}{0.8} \right\rceil \quad (2)$$

- Можно, наоборот, сформулировать требования к предельному количеству корректировок, участвующих в расчёте. Имеется в виду количество типов корректировок (на вход, на отопление и т. д.).

## Соотношение числа аналогов и корректировок III

- Максимальное количество корректировок (допускается при работе на малоактивных рынках):

$$k = \lfloor 0.8n - 3 \rfloor \quad (3)$$

- Рекомендуемое количество корректировок (используется в остальных случаях):

$$k = \lfloor \frac{0.8n - 3}{2} \rfloor \approx \lfloor -2 + 0.4n \rfloor \quad (4)$$



## Соотношение числа аналогов и корректировок IV

- Коэффициент 0.8 необходим для учёта того, что 20% аналогов не участвуют в расчёте и нужны для проверки качества модели.

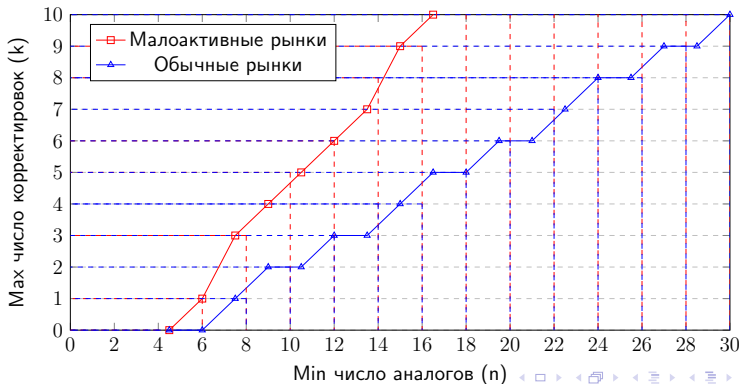
---

<sup>2</sup>Без учёта коэффициента перехода от оферты к сделке

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

**Соотношение числа аналогов и корректировок**  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Диаграмма отношения числа аналогов и корректировок



Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
**Первичный анализ данных локального конкурентного сегмента**  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Основные задачи первичного анализа I

- Первичный анализ включает в себя:
  - визуализацию распределений;
  - простейшие статистические тесты.

# Типы данных в оценке I

- Для выбора способа визуализации и теста необходимо для начала правильно классифицировать все переменные.
- Как правило оценщики имеют дело с тремя типами данных (переменных):

## Типы данных в оценке II

**Номинативные** — данные не имеющие смысл как число. Это могут быть код ОКТМО, адрес, вид разрешённого использования и т. д. Они могут быть закодированы числами, которые будут являться лишь "ярлыками". Выполнение каких-либо математических операций в отношении таких данных недопустимо. Пример: сложение или деление кодов ОКТМО абсурдно.

## Типы данных в оценке III

**Ранговые** — данные, которые можно расположить в каком-либо осмысленном порядке, так что большие значения соответствуют большему проявлению какого-либо признака по сравнению с меньшими значениями. Однако не существует какой-либо метрический аналог линейки или шкалы, чтобы определить, каково расстояние между этими категориями, или определить, одинаковы ли различия между объектами категорий 1 и 3 и

## Типы данных в оценке IV

категорий 3 и 5. Многие порядковые шкалы используются для ранжирования предпочтительности. Числа в порядковых данных несут больше смысла, чем в номинальных, и разработано много статистических методов для полного использования информации, содержащейся в упорядоченных данных, не подразумевающих ещё какие-либо свойства этих шкал. К таким данным

## Типы данных в оценке V

могут относиться "качество отделки", вид из окна, тип стен, класс энергоэффективности и т. д.

**Данные, характеризующие отношения** — характеризуются осмысленным порядком и равными интервалами между измерениями, отражающими равновеликие изменения количества любой измеренной величины, а также наличием естественного нуля. Для них являются осмысленными любые



## Типы данных в оценке VI

математические операции. К таким данным относятся: цена и стоимость, в т. ч. удельные, площадь, срок службы и т. д.

**Бинарные** — особый подтип данных, а скорее способ кодирования данных, который может быть использован для кодирования любых данных прежних типов. Это почти всегда целесообразно в случае с номинативными данными, часто целесообразно в случае с ранговыми и

## Типы данных в оценке VII

почти никогда — для данных,  
характеризующих отношения.

- Другое важное различие существует между непрерывными и дискретными данными. Непрерывные данные могут принимать любое значение вообще или в определённом диапазоне. Дискретные переменные принимают только определённые значения, и между этими значениями существуют чёткие границы.

## Типы данных в оценке VIII

- Номинативные и ранговые данные всегда дискретны. Вопрос отнесения данных, характеризующих отношения, к непрерывным или дискретным является весьма сложным и не рассматривается в данной презентации.
- При наличии достаточно большого числа наблюдений (аналогов) от 35 и выше проблема в определённой мере решается, и с определённой осторожностью можно признать допустимость обработки данных как непрерывных. Однако и при

## Типы данных в оценке IX

отсутствии такого числа аналогов всё равно приходится принимать допущение о непрерывности.

- В источниках корректировок может использоваться группировка, предполагающая дискретность самих корректировок. Использовать такие значения или аппроксимировать их до непрерывных — выбор оценщика.

# Проверка гипотез о распределении непрерывных переменных I

- Для выбора используемых в дальнейшем тестов рекомендуется выполнить проверки гипотезы о нормальном распределении для переменных в шкалах отношений.
- Для удельных цен на рынке в целом нормальность распределения является нехарактерной.
- Чаще можно говорить о приблизительно логнормальном распределении.

# Проверка гипотез о распределении непрерывных переменных II

- Однако на локальном конкурентном сегменте рынка может иметь место в т. ч. и нормальное распределение.
- Для проверки гипотезы о нормальности распределения существует достаточно много тестов.
- В частности можно предложить использование следующих тестов:
  - тест Шапиро-Уилка;
  - тест Андерсона–Дарлинга;

## Проверка гипотез о распределении непрерывных переменных III

- тест  $K^2$  Д'Агостино;
  - тест Харке Бера (один человек);
  - тест Колмогорова–Смирнова–Лиллиефорса.
- Особенностью таких тестов является то, что нулевая гипотеза говорит о нормальности распределения, т. о. маленькое р-значение (менее 0.05 либо 0.1) указывает на необходимость отклонения гипотезы о нормальности.

# Визуализация непрерывных переменных I

- Базовым способом визуализации непрерывных переменных по отдельности можно считать гистограмму. У неё существует более современный аналог — диаграмма ядерной оценки плотности.
- Недостаток гистограммы — возможность искажения содержащейся информации путём манипуляции с количеством столбцов. Следует использовать стандартные формулы определения "оптимального" числа столбцов. Либо фиксированные значения, из которых перебором выбирается наименьшее,



## Визуализация непрерывных переменных II

обеспечивающее то же субъективное восприятие, что и последующие большие значения.

- Недостаток диаграммы ядерной оценки плотности — её построение в т. ч. в области за пределами наблюдаемых значений.
- На данный момент, можно рекомендовать использование гистограммы как более простого и интуитивно понятного инструмента, при условии соблюдения правила разумно минимального количества столбцов.

## Визуализация непрерывных переменных III

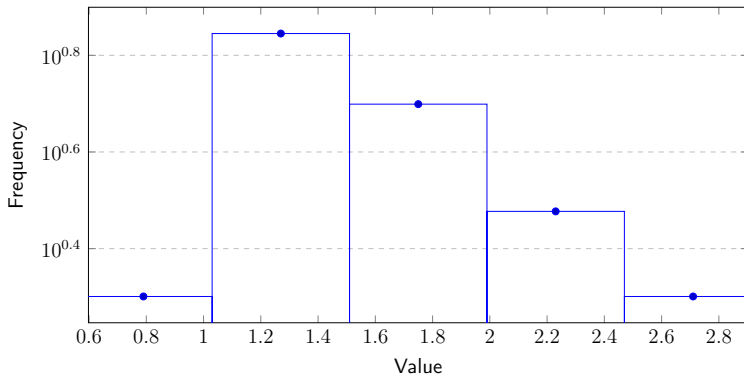
- Рекомендуется построение гистограмм как в естественной шкале, так и в логарифмах.
- Для формирования суждения оценщика о наличии связи цены (стоимости) и какой-либо непрерывной переменной полезно построение диаграмм рассеяния (scatterplot).
- Как и в случае с гистограммой полезно построение диаграмм как для естественных значений, так и для их логарифмов.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример гистограммы для 21 аналога I

Histogram of Lognormal Distribution

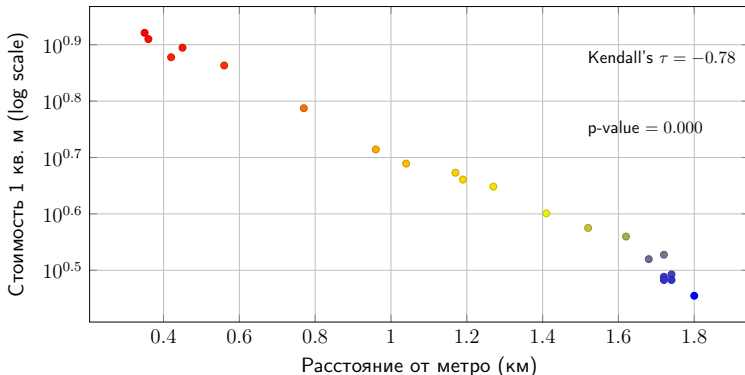


Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример диаграммы рассеяния I

Диаграмма рассеяния: удельная цена против расстояния от метро



## Выявление связи между непрерывными переменными и ценой (стоимостью) I

- Помимо собственно визуализации, необходимо показать статистическую значимость связи между переменными. Её отсутствие означает отсутствие необходимости внесения корректировки на различие в данном признаке.

## Выявление связи между непрерывными переменными и ценой (стоимостью) II

- Для упрощённого доступного начинающим оценщикам и залоговикам подхода к установлению наличия связи между признаком и стоимостью можно использовать понятие корреляции. На самом деле, всё намного сложнее. Корреляция не эквивалентна каузации, т. е. её наличие, в т. ч. статистически значимое, не означает наличие причинно-следственной связи или влияния одной переменной на другую.

## Выявление связи между непрерывными переменными и ценой (стоимостью) III

- На практике, для начинающих оценщиков сложно предложить иную простую альтернативу выявления связи между признаком и стоимостью, означающую необходимость учёта различия путём внесения корректировки.

## Выявление связи между непрерывными переменными и ценой (стоимостью) IV

- Важно знать, что нулевой гипотезой, т. е. априорным представлением о наличии связи между двумя переменными, является утверждение об отсутствии такой связи. Пока не доказано обратное, исходим из того, что такой связи нет, следовательно корректировка не нужна. Задача оценщика исследовать вопрос и попытаться опровергнуть данную нулевую гипотезу.



Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

## Выявление связи между непрерывными переменными и ценой (стоимостью) $V$

- Простым классическим средством определения наличия связи между двумя переменными в шкале отношений является расчёт коэффициента корреляции и его  $p$ -value.

## Выявление связи между непрерывными переменными и ценой (стоимостью) VI

- Первый является мерой силы связи на шкале  $[-1, 1]$ , второе — показывает статистическую значимость. Возможна ситуация, когда сам выборочный коэффициент корреляции имеет какое-то значение, но его истинное значение равно нулю, поскольку по-прежнему не возможно отклонить нулевую гипотезу об отсутствии связи.

## Выявление связи между непрерывными переменными и ценой (стоимостью) VII

- Стандартным пороговым значением  $\alpha$  для p-value в эконометрике является  $\alpha = 0.05$ . Для сверхмалых выборок ( $\approx n < 20$ ) часто целесообразно увеличить данный порог до  $\alpha = 0.1$  во избежание неконтролируемого роста вероятности ошибки второго рода, см. 1.
- Классическими мерами корреляции являются:

## Выявление связи между непрерывными переменными и ценой (стоимостью) VIII

- Коэффициент корреляции Пирсона  $r$  — параметрический метод выявления линейной связи. Он обусловлен распределением переменных и при малом их количестве (т. е. всегда в условиях оценки) требует нормальность распределения. В силу упрощённого характера метода корректировок как такового, а также недостаточной подготовки области матстатистики некоторой части оценщиков и залоговиков, его использован на практике затруднено.

## Выявление связи между непрерывными переменными и ценой (стоимостью) IX

- Коэффициент корреляции Спирмена  $\rho$  — непараметрический метод выявления линейной связи. Коэффициент инвариантен по отношению к любому монотонному преобразованию шкалы измерения. Метод является ранговым, т. е. использует не сами значения, а их ранги. Требует значительное число наблюдений, нестабилен при наличии повторяющихся значений.

## Выявление связи между непрерывными переменными и ценой (стоимостью) $X$

- Коэффициент корреляции Кенделла  $\tau$  — непараметрический метод выявления линейной связи. Коэффициент инвариантен по отношению к любому монотонному преобразованию шкалы измерения. Метод является ранговым, т. е. использует не сами значения, а их ранги. Не требует значительное число наблюдений, хорошо обрабатывает повторяющиеся значения. Рекомендуется к использования в качестве основной

# Выявление связи между непрерывными переменными и ценой (стоимостью) $X_1$

метрики связи между переменными в шкале отношений.

- Формула расчёта коэффициента корреляции Пирсона  $r$ :

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}, \quad (5)$$

где:

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

## Выявление связи между непрерывными переменными и ценой (стоимостью) XII

$x_i$  — индивидуальные значения наблюдений переменной  $x$

$y_i$  — индивидуальные значения наблюдений переменной  $y$

$\bar{x}$  — среднее по  $x$  значения

$\bar{y}$  — среднее по  $y$  значения.



## Выявление связи между непрерывными переменными и ценой (стоимостью) XIII

- Формула расчёта  $p$ -значения для коэффициента корреляции Пирсона  $r$  использует  $t$ -распределение. Для начала рассчитаем статистику  $t$ :

$$t = r \sqrt{\frac{n-2}{1-r^2}}, \quad (6)$$

## Выявление связи между непрерывными переменными и ценой (стоимостью) XIV

где:

$r$  — коэффициент корреляции Пирсона

$n$  — число наблюдений

Далее  $p$ -значение рассчитывается посредством  $t$ -распределения с  $n - 2$  степенями свободы:

$$p\text{-value} = 2 \cdot (1 - T(|t|)), \quad (7)$$

## Выявление связи между непрерывными переменными и ценой (стоимостью) XV

где  $T$  есть кумулятивная функция распределения (CDF) для t-распределения с  $n - 2$  степенями свободы.

- Формула расчёта коэффициента корреляции Пирсона  $\rho$ :

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (8)$$

где:

## Выявление связи между непрерывными переменными и ценой (стоимостью) XVI

$d_i$  — разница рангов соответствующих значений  $x_i$  и  $y_i$   
 $n$  — число наблюдений.

$r$ -значение для коэффициента корреляции Спирмена  
 $\rho$  рассчитывается аналогично тому, как для

## Выявление связи между непрерывными переменными и ценой (стоимостью) XVII

коэффициента корреляции Пирсона  $r$ : посредством  $t$ -распределения. Рассчитаем статистику  $t$ :

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}}. \quad (9)$$

Далее,  $p$ -значение рассчитывается также посредством  $t$ -распределения с  $n-2$  степенями свободы:

$$p\text{-value} = 2 \cdot (1 - T(|t|)), \quad (10)$$

## Выявление связи между непрерывными переменными и ценой (стоимостью) XVIII

где  $T$  есть кумулятивная функция распределения (CDF) для  $t$ -распределения с  $n - 2$  степенями свободы.

- Формула расчёта коэффициента корреляции Кенделла  $\tau$  без обработки повторяющихся рангов (связок):

$$\tau = \frac{2(P - Q)}{n(n - 1)}, \quad (11)$$

# Выявление связи между непрерывными переменными и ценой (стоимостью) XIX

где:

$P$  — число согласованных пар

$Q$  — число согласованных пар

$n$  — число наблюдений (объектов-аналогов)

Будем говорить, что пары  $x_i, y_i$  и  $x_j, y_j$  согласованы, если  $x_i < x_j$  и  $y_i < y_j$ , т. е.  $\text{sgn}(x_j - x_i) \text{sgn}(y_j - y_i) = 1$ .

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

## Выявление связи между непрерывными переменными и ценой (стоимостью) XX

В случае обработки повторяющихся рангов, формула выглядит следующим образом:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T_1)(P + Q + T_2)}}, \quad (12)$$



# Выявление связи между непрерывными переменными и ценой (стоимостью) XXI

где:

$P$  — число согласованных пар

$Q$  — число согласованных пар

$T_1$  — число пар, связанных только по  $x$

$T_2$  — число пар, связанных только по  $y$ .

## Выявление связи между непрерывными переменными и ценой (стоимостью) XXII

$p$ -значение для  $\tau$  Кенделла выводится из стандартного нормального распределения на основе следующей тестовой статистики  $Z$ :

$$Z = \frac{\tau \sqrt{n(n-1)(2n+5)/2}}{\sqrt{2(2n+5)}}. \quad (13)$$

Далее, рассчитаем само  $p$ -значение:

$$p\text{-value} = 2 \cdot (1 - \Phi(|Z|)), \quad (14)$$

## Выявление связи между непрерывными переменными и ценой (стоимостью) XXIII

где  $\Phi$  кумулятивная функция распределения (CDF) стандартного нормального распределения.

- Таким образом, проверяется связь каждой переменной в шкале отношений (площадь, расстояние от локального центра влияния, возраст и т. п. переменные) с показателем цен аналогов.
- Количественной мерой связи является коэффициент корреляции  $\tau$  Кенделла в совокупности с  $p$ -значением.

## Выявление связи между непрерывными переменными и ценой (стоимостью) XXIV

- Низкое значение коэффициента корреляции (см. таблицу 2) либо его статистическая незначимость (см. таблицу 3) означают отсутствие необходимости учёта различия в данном признаке.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

# Ошибки Первого и Второго рода I

	<b>Нулевая гипотеза (H0) верна</b>	<b>Нулевая гипотеза (H0) не верна</b>
<b>Отклонение H0</b>	Ошибка первого рода (False Positive)	Правильное решение (True Positive)
<b>Неотклонение H0</b>	Правильное решение (True Negative)	Ошибка второго рода (False Negative)

Таблица: Ошибки первого и второго рода

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения и

## Примерные пороговые значения для $\tau$ I

Значение $\tau$	Вывод
$ \tau  < 0.1$	Корреляция отсутствует (Negligible correlation)
$0.1 \leq  \tau  < 0.3$	Слабая корреляция (Weak correlation)
$0.3 \leq  \tau  < 0.5$	Умеренная корреляция (Moderate correlation)
$0.5 \leq  \tau $	Сильная корреляция (Strong correlation)

Таблица: Ошибки первого и второго рода

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
**Обработка данных, характеризующих отношение**  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения и

# Примерные пороговые значения для $\tau$ I

Количество аналогов	Условие значимости $\tau$
$n \leq 20$	$p < 0.1$
$n > 20$	$p < 0.05$

Таблица: Значимость р-значения

# Визуализация категориальных переменных I

- Простым способом визуализации связи цены (стоимости) с категориальными данными является диаграмма типа boxplot.
- Также можно построить совмещённые диаграммы ядерной оценки плотности.
- Ещё один способ: диаграммы средних и ошибок средних.
- Простейший вариант: stripplot.



Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

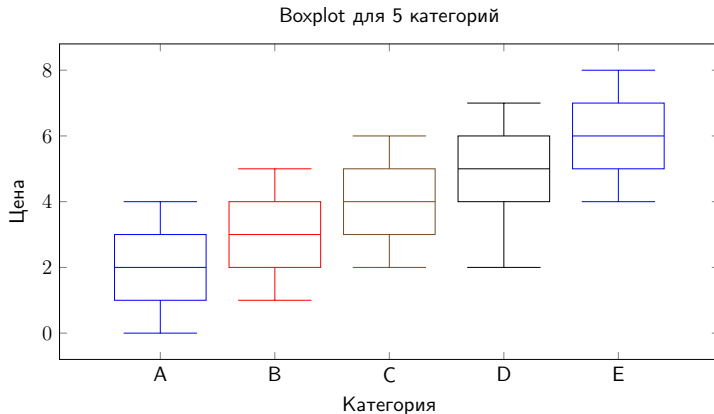
## Визуализация категориальных переменных II

- Как и в случае с данными, характеризующими отношения, полезно построение диаграмм как для естественных значений, так и для их логарифмов.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример Voxplot для 5 категорий

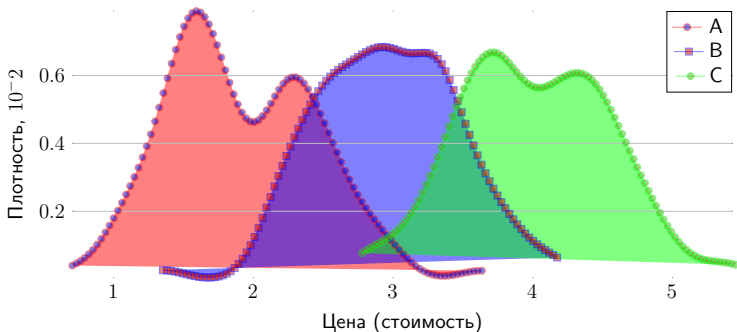


Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример диаграммы совмещённой ядерной оценки плотности для 3 категорий

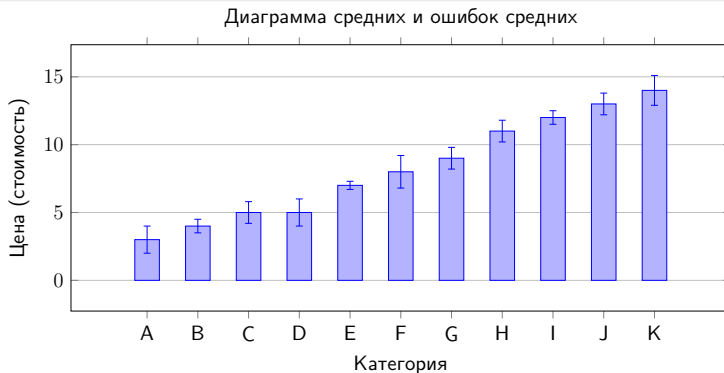
Диаграмма совмещённых ядерных оценок плотностей для 3 категорий



Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример диаграммы средних и ошибок средних

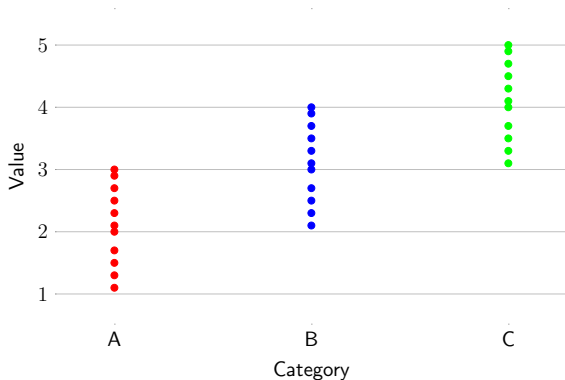


Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения к

# Пример диаграммы stripplot

Пример диаграммы stripplot



# Выявление связи между ранговыми переменными и ценой (стоимостью) I

- Для выявления наличия статистически значимых отличий показателя цены (стоимости) можно использовать два базовых классических теста: ANOVA и Критерий Краскела–Уоллиса.
- У каждого из них есть модификации и аналоги. На данном этапе можно принять решение об использовании базовых техник.

## Выявление связи между ранговыми переменными и ценой (стоимостью) II

- По умолчанию следует рекомендовать Критерий Краскела–Уоллиса, поскольку он является непараметрическим, т. е. не предъявляет требования к распределению.

## Выявление связи между ранговыми переменными и ценой (стоимостью) III

- Формула для расчёта статистики теста:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1), \quad (15)$$



## Выявление связи между ранговыми переменными и ценой (стоимостью) IV

где:

$N$  — общее количество аналогов во всех группах

$k$  — количество групп

$R_i$  — сумма рангов в  $i$ -той группе

$n_i$  количество аналогов в  $i$ -той группе.

## Выявление связи между ранговыми переменными и ценой (стоимостью) $V$

- Далее выполняется расчёт количества степеней свободы:

$$df = k - 1 \quad (16)$$

- Распределение статистики критерия согласуется с законом распределения Хи-квадрат с  $k - 1$  степенью свободы. Поэтому для расчёта  $p$ -значения можно

## Выявление связи между ранговыми переменными и ценой (стоимостью) VI

использовать кумулятивную функцию распределения (CDF) распределения  $\chi^2$ :

$$p\text{-value} = P(\chi_{df}^2 \geq H), \quad (17)$$

где:

$\chi_{df}^2$  — случайная величина, распределённая по закону  $\chi^2$  с  $df = k - 1$  степенями свободы

$H$  — расчётная статистика критерия.

## Выявление связи между ранговыми переменными и ценой (стоимостью) VII

- Расчёт p-значения:

$$p\text{-value} = 1 - F(H \mid \chi_{k-1}^2). \quad (18)$$

- Его интерпретация аналогична той, что показана в таблице 3.
- Можно использовать ANOVA для логарифмов цен (стоимостей).

## Выявление связи между ранговыми переменными и ценой (стоимостью) VIII

- На первом шаге считаем среднее для всех аналогов и среднее для каждой группы.
- Далее считаем межгрупповую изменчивость:

$$SSB = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2, \quad (19)$$

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

## Выявление связи между ранговыми переменными и ценой (стоимостью) IX

где:

$n_i$  — количество аналогов в группе  $i$

$\bar{X}_i$  — среднее значение логарифма цены в группе  $i$

$\bar{X}$  — среднее значение логарифма цены всех аналогов.

## Выявление связи между ранговыми переменными и ценой (стоимостью) $X$

- Далее считаем внутригрупповую изменчивость:

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \quad (20)$$

где:

$X_{ij}$  —  $j$ -й аналог в группе  $i$ .

## Выявление связи между ранговыми переменными и ценой (стоимостью) $X_i$

- Расчёт F-статистики:

$$F = \frac{MSB}{MSW}, \quad (21)$$

где:

$$MSB = \frac{SSB}{k - 1} \quad (22)$$

и

$$MSW = \frac{SSW}{N - k}, \quad (23)$$



## Выявление связи между ранговыми переменными и ценой (стоимостью) XII

где:

$k$  — количество групп

$N$  — общее число аналогов.

- Рассчитаем количество степеней свободы:

$$df_{MSB} = k - 1. \quad (24)$$

$$df_{MSW} = N - k. \quad (25)$$

## Выявление связи между ранговыми переменными и ценой (стоимостью) XIII

- Значение  $p$ -value рассчитывается с помощью кумулятивной функции распределения (CDF)  $F$ -распределения. Это вероятность того, что  $F$ -статистика будет такой же или более экстремальной, чем значение, рассчитанное по данным выборки, при нулевой гипотезе о том, что все групповые средние значения равны.

## Выявление связи между ранговыми переменными и ценой (стоимостью) XIV

- Формула.

$$p\text{-value} = P(F_{k-1, N-k} \geq F), \quad (26)$$

где:

$(F_{k-1, N-k}$  — F-распределение с  $k - 1$  и  $N - k$  степенями свободы.

$F$  — расчётное значение F-статистики.

## Выявление связи между ранговыми переменными и ценой (стоимостью) XV

- Более понятный вид формулы:

$$p\text{-value} = 1 - F(F | F_{k-1, N-k}), \quad (27)$$

где:

$F$  — расчётное значение F-статистики  
( $F_{k-1, N-k}$  — F-распределение с  $k - 1$  и  
 $N - k$  степенями свободы)

## Выявление связи между ранговыми переменными и ценой (стоимостью) XVI

- Интерпретация  $p$ -значения аналогична той, что показана в таблице 3.
- Оценщик сам вправе выбирать метод проверки наличия значимости. Рекомендуются использование обоих методов сразу.

## Выявление связи между ранговыми переменными и ценой (стоимостью) XVII

- Напоминаем, что нулевой гипотезой является отсутствие различий между группами, следовательно отсутствие необходимости внесения корректировки. Необходимость её применения должна быть доказана, но не наоборот.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

# Нулевая и альтернативная гипотеза для ANOVA и Критерия Краскелла–Уоллиса I

Тест	Нулевая гипотеза $H_0$	Альтернативная гипотеза $H_1$
Предпосылки	Не требует доказательства, верна по умолчанию	Требует доказательство
ANOVA	Средние значения логарифмов цен равны во всех группах	Средние значения логарифмов цен отличаются как минимум между двумя группами
Краскела–Уоллиса	Медианы цен равны во всех группах	Медианы цен отличаются как минимум между двумя группами

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
**Обработка категориальных переменных**  
Обработка бинарных данных  
Формирование итогового суждения о необходимости применения

# Нулевая и альтернативная гипотеза для ANOVA и Критерия Краскелла–Уоллиса II

**Таблица:** Нулевая и альтернативная гипотезы для ANOVA и Критерия Краскела–Уоллиса



# Визуализация бинарных данных I

- Для визуализации бинарных данных с точки зрения их влияния на стоимость подходят те же методы, что и для категориальных.
- Во всех случаях будут две категории.

# Выявление связи между бинарными переменными и ценой (стоимостью) I

- Здесь также можно предложить два теста: один непараметрический, второй параметрический. Непараметрический следует применять к самим данным о ценах, параметрический к их логарифмам.
- Есть немало различных тестов. Для начала можно предложить два классических.

## Выявление связи между бинарными переменными и ценой (стоимостью) II

- Непараметрический критерий: тест Манна–Уитни–Уилкоксона (в народе U-тест). Есть несколько трактовок того, какую именно гипотезу он проверяет:
  - гипотеза о равенстве распределений;
  - гипотеза сдвига;

## Выявление связи между бинарными переменными и ценой (стоимостью) III

- гипотеза стохастического превышения значения признака (в нашем случае стоимости, скорее всего удельной) объектов, принадлежащих к одной группе относительно объектов, принадлежащих другой группе.

## Выявление связи между бинарными переменными и ценой (стоимостью) IV

- В целом, это не так важно. Важно: нулевая гипотеза по-прежнему говорит о том, что между двумя группами нет различий в стоимости, следовательно внесение корректировки на различие в разделяющем признаке не требуется. Оценщик может попытаться доказать наличие различия и необходимость внесения корректировки. Но не наоборот.

## Выявление связи между бинарными переменными и ценой (стоимостью) $V$

- Параметрическим критерием является тест Стьюдента для независимых выборок (в народе  $t$ -тест). Его следует применять к логарифмам цена аналогов.
- Он проверят равенство средних двух групп.
- Предположим, нам нужно установить значимость влияния отдельного входа на стоимость. Изначально предполагаем, что влияния нет.
- Начнём с  $U$ -теста.

## Выявление связи между бинарными переменными и ценой (стоимостью) VI

- Объединим две выборки в одну и проранжируем в возрастающем порядке показателя стоимости. В случае наличия повторяющихся значений, берём средний ранг по ним.
- $n_1$  и  $n_2$  — количество наблюдений в каждой из групп.  
 $R_1$  и  $R_2$  — сумма рангов в каждой из групп.

## Выявление связи между бинарными переменными и ценой (стоимостью) VII

- Рассчитаем значения U-статистики для обеих групп:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$
$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2.$$
(28)

В качестве U-статистики принимается меньшее значение:

$$U = \min(U_1, U_2).$$
(29)



## Выявление связи между бинарными переменными и ценой (стоимостью) VIII

Распределение U-статистики может быть аппроксимировано нормальным распределением, со средним и стандартным отклонением, определяемыми как:

$$\begin{aligned}\mu_U &= \frac{n_1 n_2}{2} \\ \sigma_U &= \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}\end{aligned}\tag{30}$$

соответственно.

## Выявление связи между бинарными переменными и ценой (стоимостью) IX

- Далее идёт вычисление z-метки:

$$z = \frac{U - \mu_U}{\sigma_U}. \quad (31)$$

Само p-значение вычисляется из нормального распределения:

$$p = 2(1 - \Phi(|z|)) \quad (32)$$

## Выявление связи между бинарными переменными и ценой (стоимостью) $X$

- Интерпретация результатов аналогична всем предыдущим, см. таблицу 3.
- В случае возможности применения параметрических методов, также целесообразно провести t-тест для независимых выборок.

## Выявление связи между бинарными переменными и ценой (стоимостью) $X_1$

- Выборочные средние и дисперсии определены как:

$$\bar{X}_1, \bar{X}_2 \\ s_1^2, s_2^2$$

соответственно.

- Рассчитаем совокупное стандартное отклонение:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (33)$$

## Выявление связи между бинарными переменными и ценой (стоимостью) XII

- Далее, необходимо провести расчёт t-статистики:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \quad (34)$$

- Определение числа степеней свобод:

$$df = n_1 + n_2 - 2 \quad (35)$$

## Выявление связи между бинарными переменными и ценой (стоимостью) XIII

- $p$ -значение рассчитывается из  $t$ -распределения:

$$p = 2(1 - T(|t|, df)) \quad (36)$$

- Интерпретация результатов аналогична всем предыдущим, см. таблицу 3.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
**Формирование итогового суждения о необходимости применения и**

# Формирование итогового суждения о необходимости применения конкретных корректировок I

- Оценщик выявляет переменные (признаки), оказывающие существенное влияние на стоимость с помощью показанных выше методов либо иных аналогичных.

## Формирование итогового суждения о необходимости применения конкретных корректировок II

- К каждому аналогу можно применять не более 2 корректировок. Метод корректировок в принципе не является научным, поэтому невозможно строго доказать конкретное требование. Однако есть ряд соображений.



## Формирование итогового суждения о необходимости применения конкретных корректировок III

- Исходим из того, что метод корректировок это всё же не случайное блуждание по ценам аналогов. Оценщик формирует модель ценообразования, пускай и достаточно слабую.

## Формирование итогового суждения о необходимости применения конкретных корректировок IV

- В случае наличия двух корректировок и применения мультипликативной модели, определение скорректированной стоимости выглядит так:

$$Y'_i = Y_i(a1 \times a2), \quad (37)$$

# Формирование итогового суждения о необходимости применения конкретных корректировок $V$

где:

$Y_i$  — наблюдаемая цена аналога

$Y'_i$  — скорректированная цена аналога

$a_1$  — корректировка 1

$a_2$  — корректировка 2

## Формирование итогового суждения о необходимости применения конкретных корректировок VI

Зная обе цены и значение одной из корректировок, мы всегда можем найти значение корректировки, являющееся единственно возможным. Например:

$$Y'_i = Y_i(a_1 \times a_2)$$
$$a_2 = \frac{Y'_i}{Y_i \times a_1} \quad (38)$$

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
**Формирование итогового суждения о необходимости применения и**

## Формирование итогового суждения о необходимости применения конкретных корректировок VII

- В то же время, наличие хотя бы трёх корректировок, даст бесконечное число вариантов для  $a_2$  и  $a_3$ .
- Таким образом, для обеспечения относительной стабильности модели и устойчивости влияние переменных, необходимо применение не более двух корректировок к каждому аналогу.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
**Теоретические требования к расчётам**  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Соотношение числа аналогов и корректировок  
Первичный анализ данных локального конкурентного сегмента  
Типы данных в оценке  
Проверка гипотез о распределении непрерывных переменных  
Обработка данных, характеризующих отношение  
Обработка категориальных переменных  
Обработка бинарных данных  
**Формирование итогового суждения о необходимости применения и**

## Формирование итогового суждения о необходимости применения конкретных корректировок VIII

- Выбор аналогов и их соответствие данным критериям является обязанностью оценщика.

# Визуальный анализ I

- Поскольку задачей метода корректировок является устранение различий цены, вызванных различием в признаках, внесение корректировок на эти признаки должно приводить к устранению этого различия.
- Это можно визуализировать путём сравнения распределения цен до внесения корректировок и после него. В идеальном случае, распределение становится равномерным. В типичном, происходит сглаживание. В плохом: распределение как был

## Визуальный анализ II

переворачивается, и его асимметрия меняет направление.

- Следующий этап: визуализация изменения связи между непрерывными переменными. В идеальном случае связь полностью исчезает: все точки лежат на одной прямой параллельно X. В типичном, есть явное уменьшение наклона. Недопустимый вариант: изменение наклона.



## Визуальный анализ III

- Далее следует анализ для категориальных и бинарных переменных. Влияние корректировок на различия в них в идеале приводит к полному устранению влияния значений на скорректированные цены. На практике, чаще всего — частичному. Недопустимый случай: переверот влияния: т. е. перестановка категорий.
- Выполнение такой визуализации следует делать как для анализа влияния каждой корректировки, так и по итогам внесения всех корректировок.

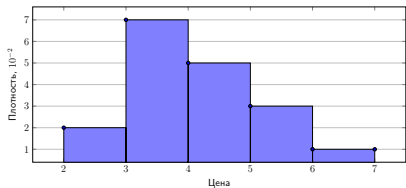
- Используемые обозначения
- История разработки
- Постановка задачи
- Общее описание метода корректировок
- Основные проблемы расчётов
- Теоретические требования к расчётам
- Методы и метрики проверки готовых расчётов
- Методы валидации и тестирования
- Практическая реализация на уровне MVP

### Визуальный анализ

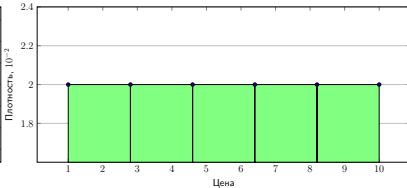
- Проверка изменения распределения цен аналогов
- Проверка устранения связи между ценой и переменными

# Визуализация изменения распределения

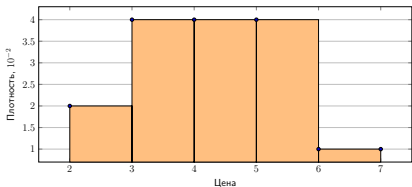
Initial (Lognormal)



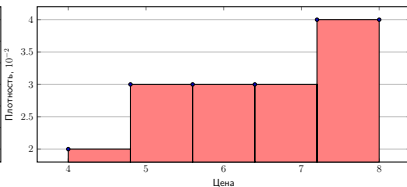
Perfect (Uniform)



Normal (Gentle Normal)



Bad (Peaked Right)



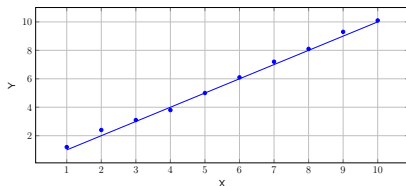
Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

## Визуальный анализ

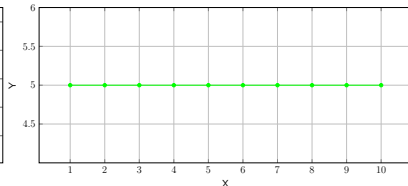
Проверка изменения распределения цен аналогов  
Проверка устранения связи между ценой и переменными

# Визуализация изменения корреляции

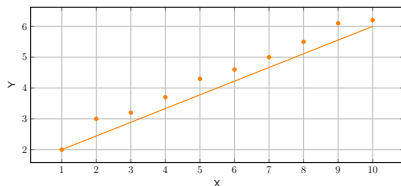
Initial (Significant Slope)



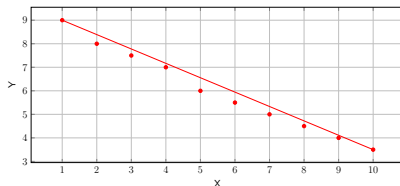
Ideal (Straight Line)



Typical (Reduced Slope)



Bad (Changed Slope Direction)

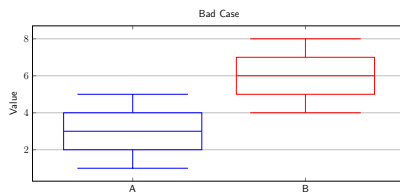
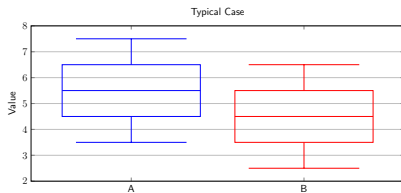
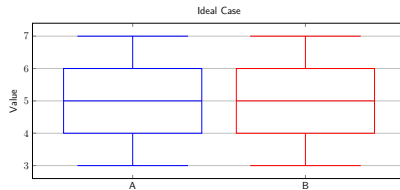
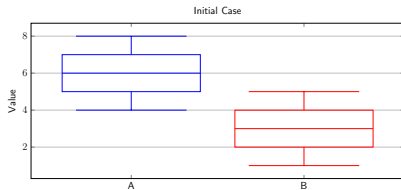


Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

## Визуальный анализ

Проверка изменения распределения цен аналогов  
Проверка устранения связи между ценой и переменными

# Визуализация изменения распределения



# Изменение распределения, тест на равномерное распределение I

- Поскольку задача метода корректировок это уменьшение различия между аналогами и объектом оценки, а также между аналогами, итогом внесение корректировок должно стать равномерное распределение цен аналогов.
- Для проверки гипотезы о равномерности распределения скорректированных цен можно использовать следующие тесты:
  - тест  $\chi^2$ ;

# Изменение распределения, тест на равномерное распределение II

- тест Колмогорова–Смирнова;
- тест Андерсона–Дарлинга.
- Рассмотрим тест  $\chi^2$ . Его нулевая гипотеза гласит, что данные распределены равномерно.
- Для начала необходимо разбить данные на  $k$  равных интервала.

## Изменение распределения, тест на равномерное распределение III

- Пусть  $O_i$  это эмпирические частоты для интервала  $i$ ,  
 $O_i$  это теоретические частоты для интервала  $i$ ,  
определяемые как:

$$E_i = \frac{n}{k}, \quad (39)$$

где  $n$  — общее количество наблюдений.

## Изменение распределения, тест на равномерное распределение IV

- Рассчитаем статистику  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}. \quad (40)$$

- $p$ -значение вычисляется из самого распределения  $\chi^2$  с  $k - 1$  степенями свободы  $p = P(\chi^2 \geq \chi_{\text{observed}}^2 | k - 1)$ :

$$p = 1 - \Gamma\left(\frac{k-1}{2}, \frac{\chi^2}{2}\right). \quad (41)$$



# Изменение распределения, тест на равномерное распределение $V$

- Рассмотрим тест Колмогорова-Смирнова. Он имеет аналогичную нулевую гипотезу.
- Рассчитаем функцию эмпирического распределения (EDF). EDF,  $F_n(x)$  доля аналогов, имеющих значение меньше либо равно  $x$ .

# Изменение распределения, тест на равномерное распределение VI

- Рассчитаем статистику теста:

$$D = \sup_x |F_n(x) - F(x)|, \quad (42)$$

где:

$F_n(x)$  — эмпирическая функция распределения

$F(x)$  — кумулятивная функция распределения для равномерного

# Изменение распределения, тест на равномерное распределение VII

- Для непрерывного распределения:

$$F(x) = x \quad \text{for } x \in [0, 1]. \quad (43)$$

- р-значение статистики D вычисляется для нулевой гипотезы:

$$p = P(D \geq \text{observed value}) \quad (44)$$

# Изменение распределения, тест на равномерное распределение VIII

вычисляется с помощью аппроксимации распределением Колмогорова:

$$p = 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2 D^2 n), \quad (45)$$

где  $n$  — количество наблюдений.

# Изменение распределения, тест на равномерное распределение IX

- Тест Андерсона-Дарлинга основан на тех же принципах, что и Колмогорова-Смирноа, и придаёт больший вес хвостам распределения. Он предполагает аналогичную нулевую гипотезу.
- Предположим, что  $F(x)$  является CDF для равномерного распределения.

# Изменение распределения, тест на равномерное распределение $X$

- Статистика теста  $A^2$  вычисляется следующим образом:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln(F(X_i)) + \ln(1 - F(X_{n+1-i}))] \quad (46)$$

# Изменение распределения, тест на равномерное распределение $X_1$

- Вычисление р-значения осуществляется для нулевой гипотезы  $p = P(A^2 \geq \text{observed value})$  следующим образом:

$$A_{\text{adjusted}}^2 = A^2 \left( 1 + \frac{0.75}{n} + \frac{2.25}{n^2} \right) \quad (47)$$

# Изменение распределения, тест на равномерное распределение $\chi^2$

$$p = \begin{cases} 1 - \exp\left(-13.436 + 101.14A_{\text{adjusted}}^2 - 223.73(A_{\text{adjusted}}^2)^2\right) & \text{if } A_{\text{adjusted}}^2 < 0.2 \\ 1 - \exp\left(-8.318 + 42.796A_{\text{adjusted}}^2 - 59.938(A_{\text{adjusted}}^2)^2\right) & \text{if } A_{\text{adjusted}}^2 < 0.34 \\ \exp\left(0.9177 - 4.279A_{\text{adjusted}}^2 - 1.38(A_{\text{adjusted}}^2)^2\right) & \text{if } A_{\text{adjusted}}^2 < 0.6 \\ \exp\left(1.2937 - 5.709A_{\text{adjusted}}^2 + 0.0186(A_{\text{adjusted}}^2)^2\right) & \text{otherwise} \end{cases} \quad (48)$$

- В случае неотклонения гипотезы о равномерности распределения, можно сделать вывод о выполнении задач корректировок в этой части.



# Проверка устранения связи между ценой и переменными I

- Теперь перед нами стоит задача противоположная той, что была ранее.
- Нам необходимо убедиться в том, что скорректированные цены не имеют значимую корреляцию с непрерывными переменными. Т.е. коэффициент корреляции  $\tau$  должен иметь незначимое р-значение либо, как минимум существенно снизиться (в 2.618 и более раз). Не допускается смена знака.

# Проверка устранения связи между ценой и переменными II

- Также проверяется отсутствие связи скорректированных цен с категориальными и бинарными переменными. Также недопустима смена направления влияния. Поэтому выполняются в т. ч. односторонние тесты.
- Проверка выполняется как по каждой корректировке отдельно, так и в отношении итоговых скорректированных цен.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Визуальный анализ  
Проверка изменения распределения цен аналогов  
Проверка устранения связи между ценой и переменными

# Проверка устранения связи между ценой и переменными III

- В случае невыполнения данных условий, расчёт подлежит пересмотру.

# Методы валидации I

- Для выполнения валидации и тестирования, оценщик предоставляет расчёт в формате электронной таблицы. Далее, он сам либо сотрудник залоговой службы создаёт из неё три файла CSV. Об этом подробнее в следующей части.

## Методы валидации II

- Задача валидации есть оценка качества модели ценообразования, построенной с помощью 80% аналогов. Система делит аналоги на обучающую и тестовую подвыборки в пропорции 80:20 псевдослучайным образом. Обозначим их как  $S$  и  $T$  соответственно.

## Методы валидации III

- Существует:

$$\binom{C}{T} \quad (49)$$

способов разделения аналогов на расчётную и тестовую выборки. Рассмотрим пример.

Предположим, у нас есть 24 аналога. В этом случае в расчётную выборку идут 19 аналогов, в тестовую — 5. Тогда:

$$\binom{24}{5} = \frac{24!}{5!(24-5)!} = \frac{24!}{5! \times 19!} = \frac{5100480}{120} = 42504$$

## Методы валидации IV

Таким образом, существует 42504 уникальных способов разделения 24 аналогов в пропорции 19:5.

- Следующий этап валидации: последовательное исключение одного аналога из расчёта и его оценка на основе оставшихся аналогов. В рассматриваемом случае, будет проведено 19 валидационных расчётов. Проводится расчёт метрик средней абсолютной и средней квадратичной ошибок (MAD и MSE соответственно).

## Методы валидации V

- Далее проводится аналогичная валидация при удалении двух аналогов.

$$\binom{19}{2} = \frac{19!}{2!(19-2)!} = \frac{19 \times 18}{2 \times 1} = 171 \quad (50)$$

- Затем аналогично для трёх аналогов.

$$\binom{19}{3} = \frac{19!}{3!(19-3)!} = \frac{19 \times 18 \times 17}{3 \times 2 \times 1} = 969 \quad (51)$$



## Методы валидации VI

- Таким образом, всего выполняется 1149 итераций валидации. Она помогает решать многие задачи, в т. ч. выявлять аномальные аналоги.
- Валидация сопровождается визуализациями распределения ошибок, расчётами метрик ошибок, например критерия Дарбина-Уотсона:

$$D = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (52)$$

## Методы валидации VII

Ошибки должны иметь нормальное распределение со средним в окрестностях нуля. У них не должно быть автокорреляции. В целом, можно подсмотреть идеи из регрессионного анализа и применять принятые методы анализа остатков.

- Ylisjärvi использует 8 критериев анализа валидационных расчётов одного аналога на основе остальных.

## Методы валидации VIII

- Дополнительно проводится расчёт всех аналогов без одной из корректировок с анализом значимости изменений результатов с помощью t- и U-тестов. В случае незначительности изменения, данную корректировку следует удалить из расчётов.
- Итогом валидации является суждение о качестве модели и её применимости либо неприменимости для выполнения расчёта объекта оценки. Оценщику следует помнить, что он несёт ответственность, в т. ч. имущественную за результаты своей работы.

## Методы валидации IX

Поэтому не следует торопиться с принятием решения о готовности модели.

- В случае неудовлетворительной работы модели, следует пересмотреть состав и конкретные значения корректировок, заменить часть аналогов либо увеличить их количество. После этого строится новая модель с повторением всех шагов.

# Итоговое тестирование I

- *Практика есть критерий истины* — Карл Маркс.
- Существует большое число теорем, методов. Определить, какой метод лучше в данном случае на основе рассуждений — невозможно. Нужно применить его и показать качество работы на данных, на которых он не обучался. А затем сравнить с результатами других методов.

## Итоговое тестирование II

- В данных по аналогам содержатся "сигнал" и "шум". Любая модель обучается, как на сигнале, так и на шуме. Пример: полиномиальная регрессия степени  $n - 1$ , где  $n$  — количество аналогов, даст иллюзию идеальной модели оценки. На самом деле, она бесполезна: она знает всё о наборе аналогов и ничего об окружающем мире.
- Поэтому тестирование проводится строго на обособленных данных.

## Итоговое тестирование III

- Задача тестирования есть итоговая оценка качества модели ценообразования, созданной посредством внесения корректировок.
- Аналоги, попавшие в тестовую выборку, никак не участвуют в дообучении модели корректировок.
- Для тестирования следует использовать 20 % аналогов, не участвовавших в расчёте.
- Тестирование проводится путём оценки этих аналогов по той модели, которая была получена в ходе выбора и применения расчётов и валидации.

## Итоговое тестирование IV

- Как правило, единственной метрикой в данном случае является то или иное значение ошибки: MAD и (или) MSE.
- Данная метрика указывает на ожидаемую погрешность оценки стоимости и может служить метрикой для расчёта вероятного диапазона значения стоимости. Т.е. оценщик получает доказанную меру неопределённости, способную, в т. ч. дать ему аргументы для своей защиты в случае каких-либо претензий.



## Итоговое тестирование V

- Итоговое суждение о применимости конкретной модели корректировок выносится на основе значения ошибки. Для целей описанной разработки, допустимой ошибкой на тестовой выборке считались:

$$|error| = \begin{cases} \sqrt[4]{1.618} - 1 = 0.128 & \text{активные рынки} \\ \sqrt{1.618} - 1 = 0.272 & \text{неактивные рынки.} \end{cases} \quad (53)$$

# Пройденный путь I

- Для выполнения всего, описанного выше, и ряда иных дополнительных тестов был разработан модуль *Ylisjärvi*. Фактически до минимально работоспособного уровня MVP (Minimum Viable Product) был доведён только данный back-end.
- К нему должен был идти front-end *Suursaari*, предоставляющий возможность загрузки файла с расчётом в обычном браузере в локальной сети заказчика и получение результатов проверки в виде HTML либо PDF файла.

## Пройденный путь II

- Ставился, но был рассмотрен вопрос о возможности доступа к системе для оценщиков также через браузер с целью обеспечения возможности выполнять ими проверку самостоятельно для ускорения процедуры выдачи займа, а также постепенного формирования пула грамотных специалистов, способных разбираться в относительно новых для них вопросах анализа данных и доводить свои расчёты для относительно приемлемого уровня.

## Пройденный путь III

- Как было сказано в начале, проект был остановлен ввиду решения использовать ещё более простой подход: проводить оценку силами ML-модели.
- Для выполнения проверки необходимо сначала сформировать файлы по определённому шаблону. Для этого оценщик передаёт свой расчётный файл, сотрудник залоговой службы по инструкции с помощью набора правил создаёт 3 новых листа, содержащих данные в определённом формате с определённой разметкой:
  - лист с данными по всем аналогам;

## Пройденный путь IV

- лист с данными по объекту или объектам оценки;
- лист с набором правил для корректировок.
- Далее, листы сохраняются как отдельные файлы в формате CSV.
- Файлы загружаются через интерфейс.
- Дальнейшее участие человека не требуется. Вся проверка идёт автоматически. Был поставлен, но не рассматривался вопрос о возможности задания некоторых настроек вручную. В базовом варианте *Ylisjärvi* принимает все решения автоматически.

## Пройденный путь V

- Результатом является отчёт о проверке в формате HTML или PDF, содержащий значительное количество метрик, диаграмм, часть из которых описана в презентации. Также присутствует вывод о допустимости либо недопустимости расчёта в качестве основания для принятия решения о стоимости залога. Содержатся некоторые рекомендации.
- Предполагалась работа в двух режимах:
  - автономная проверка;

# Пройденный путь VI

- полная проверка.
- В первом режиме выполняется то, что показано в презентации и ряд иных тестов.
- Во втором проводится анализ расчёта на соответствие рыночным условиям.
- В отношении объекта оценки выполняются в т. ч. следующие вычисления:

## Пройденный путь VII

- для всех признаков в непрерывных шкалах вычисляются квантили (например, можно узнать, что 55% объектов на рынке региона имеют меньшую площадь, чем объект оценки, что указывает на его типичность);
- также вычисляется его расстояние от среднего значения в логарифмах в шкале стандартных отклонений, например можно узнать, что объект оценки находится на расстоянии в  $+0.57$  стандартных отклонений от среднего, что также указывает на его типичность.



## Пройденный путь VIII

- для категориальных и бинарных переменных вычисляется доля объектов в каждой группе, после чего объект оценки попадает в одну из них, что может, например, дать знание о том, что он имеет такое же состояние отделки, как 37% объектов на рынке, что также делает его типичным.
- В отношении объектов аналогов проводятся такие же тесты, а также тесты для показателей стоимости: абсолютной и удельной.

## Пройденный путь IX

- Иногда это даёт сделать определённые интересные выводы: например, объект оценки по всем своим характеристикам относится к "типичным", а объекты-аналоги нет. Например, их квантили удельной стоимости находятся на отрезке  $[0.85-0.98]$ , а их расстояние от среднего значения от логарифма удельной стоимости в стандартных отклонения —  $[+2.45-+2.93]$  что указывает на тенденцию брать дорогие объекты при наличии более дешёвых и при этом более схожих с объектом оценки, являющимся "типичным" объектом на своём рынке. При этом,

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Пройденный путь  
Печальный финал  
Новые возможности

## Пройденный путь X

формальные критерии расчёта могут быть соблюдены, и он может пройти проверку в автономном режиме.

# Печальный финал I

- Необходимость заблаговременной разработки противооценщиковой антифрод системы, заведомо опережающей растущие возможности оценщиков по манипуляции расчётами с целью преодоления системы контроля, описанной в данной презентации, привела к сбору значительных данных по рынкам регионов деятельности, что натолкнуло на идею о возможности ввести третью систему для борьбы с недобросовестными оценщиками: самостоятельный расчёт с помощью скоринговой модели, обученной

## Печальный финал II

на данных регионального рынка, с последующей сверкой результатов данной системы и у оценщика. Если результат его оценки ниже результата модели, то его отчёт принимается без проверки для формальных целей.

- На практике, во время тестирования отдельных расчётов, примерно 80–90 % проверок оканчивались неуспешно из-за неспособности залоговиков правильно разметить данные в файле. 10–20 % проверок были провалены на первом этапе проверки

## Печальный финал III

количества аналогов и корректировок. Типичная ситуация: от 3 до 5 аналогов и от 5 до 8 корректировок.

- Всё это стало частью причин отказа от продолжения разработки в пользу решения на основе XGBoost, в принципе не требующего участие человека в процессе определения стоимости.

Используемые обозначения  
История разработки  
Постановка задачи  
Общее описание метода корректировок  
Основные проблемы расчётов  
Теоретические требования к расчётам  
Методы и метрики проверки готовых расчётов  
Методы валидации и тестирования  
Практическая реализация на уровне MVP

Пройденный путь  
Печальный финал  
Новые возможности

## Печальный финал IV

- Если система должна быть на порядок умнее оценщика и залоговика, чтобы не дать им навредить умышленно или по незнанию, зачем вообще нужны оценщик и проверяющий?

# Новые возможности I

- Несмотря на неудачный опыт внедрения данной системы проверки корректировочных таблиц, сделанные разработки могут представлять определённый интерес для самих оценщиков, СРО и проверяющих из менее крупных структур.
- Сформулированы общие требования, полный перечень тестов и критериев проверки может быть свободно опубликован и использоваться для повышения качества расчётов методом корректировок.



## Новые возможности II

- Процесс доработки бекэнда и разработки фронтенда замедлился, но продолжается. Ориентировочный срок запуска: IV квартал 2025 года – II квартал 2026 года. Вероятно, это будет бесплатный сервис.
- Процесс внедрения математических моделей, усиленных возможностями оборудования, будет продолжаться дальше. Некоторые авторы считают, что проблемой таких методов является сложная интерпретация результатов. Тем не менее, существуют легко интерпретируемые модели машинного обучения.

## Новые возможности III

- Для предоставления понятного руководству либо заказчику и всем заинтересованным пользователям услуг по оценке результата, перспективным направлением может являться применение интрепретируемых нейронных сетей на основе теоремы Колмогорова–Арнольда, содержащей решение 13-й проблемы Гильберта. В данной теореме заложена основа представления непрерывных функций нескольких переменных

## Новые возможности IV

суперпозициями непрерывных функций меньшего числа переменных.

$$f(x_1, x_2, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left( \sum_{p=1}^n \psi_{qp}(x_p) \right), \quad (54)$$

где:  $\Phi_q$  и  $\psi_{qp}$  непрерывные функции  $[0, 1] \rightarrow \mathbb{R}$ .

Функции  $\Phi_q$  зависят только от одной переменной, а  $\psi_{qp}$  — от  $x_p$ .

- Об интерпретируемых нейронных сетях в оценке когда-нибудь в другой раз. Спасибо за внимание!