

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет экономических наук

Проектная работа по дисциплине "Эконометрика 2" на тему:

**"Оценивание гедонистической ценовой функции для
фотоаппаратов с помощью моделей множественной регрессии"**

ФИО	Вклад	Процент вклада
Амиров Марат Наильевич	Сбор данных, первичный анализ, квантильная регрессия	50
Оськин Егор Дмитриевич	Сбор данных, проверка ТГМ, выбор модели,	50

Руководители: Вакуленко Е.С., Погорелова П.В.

Москва 2024

1 Введение и постановка задачи

(Для рецензирующих и проверяющих - просьба идти по `irunb` файлу в хронологическом порядке, так будет понятен порядок приложений и результатов, очень много информации есть содержится только там).

Мы поставили перед собой задачу — понять, насколько некоторые технические и маркетинговые характеристики фотокамер влияют на их стоимость, а также построить модель, которая могла бы предсказывать стоимость фотоаппарата (фотокамеры, далее — синонимы) по заданным меркам.

В своём исследовании данные и характеристики камер мы собирали вручную с сайта: <https://versus.com/ru/camera>. Все фотокамеры в нашем исследовании цифровые и выпущены после 2012 года, чтобы сохранить связь между ценами и характеристиками (мы постарались исключить коллекционные и редкие модели, чтобы избежать выбросов).

2 Переменные

2.1 Зависимая переменная

Цену фотокамер (в данных — “Цена”), мы собирали двух ресурсов: Яндекс.Маркет и Озон, цены учитывались только у проверенных продавцов без учёта скидок маркетплейсов. В случае отсутствия данных в российском сегменте интернета мы собирали цены через Cdek.Shopping или Amazon (с дальнейшим переводом цен из валюты по текущему курсу). В случае сильного расхождения цены в наших источниках (более 5%), мы брали среднюю цену на продукт. Также стоит сказать, что все цены были посчитаны на момент 22.05.2025 и не менялись с момента агрегации.

2.2 Регрессоры

Из-за большого количества характеристик фотоаппаратов, влияющих на цену, мы разделили их на классы по функциональному назначению и выбрали наиболее релевантные признаки:

Дизайн и бренд-характеристики

Производитель — категориальная переменная, определяющая марку фотоаппарата. Значения: Nikon, Leica, Canon, Sony, Fujifilm, Panasonic).

Вес (в килограммах) — непрерывная переменная, отражающая размеры и эргоно-

мичность фотоаппарата.

Наличие пыли/ влагозащиты — бинарная переменная.

Оптические характеристики

Мегапиксели (основная камера) — непрерывная переменная, отражающая разрешение камеры.

Максимальный ISO — непрерывная переменная, показывающая светочувствительность камеры.

Максимальный расширенный ISO — непрерывная переменная, увеличивающая светочувствительность через программные средства.

Непрерывная съёмка FPS — непрерывная переменная, показывающая скорость затвора камеры. При отсутствии функции серийной съёмки присваивалось значение 1.

Видео характеристики

Видео (качество) — категориальная переменная, показывающая максимальное качество видео.

Видеозапись fps — категориальная переменная, указывающая максимальное количество кадров в секунду в видео: 24, 30, 60, 120.

Показатели аккумулятора

Время работы батареи (shots) — дискретная переменная, указывающая количество кадров на одном заряде.

Ёмкость аккумулятора — непрерывная переменная, показывающая общий энергозаряд.

3 Первичный анализ данных

Многие из наших переменных — категориальные, поэтому описательные характеристики для них малоинформативны, в дальнейшем мы непосредственно посмотрим на их распределения в нашей выборке. Дополнительно мы посмотрели на квартили и интерквартильный размах, так как планируем использовать квантильную регрессию в дальнейшем.

Построив ковариационную матрицу для всех регрессоров, мы заметили, что у нас нет причин полагать, что между признаками наблюдается сильная мультиколлинеарность. Самое высокое значение связи наблюдается между признаками “Наличие пыли/ влаго защиты” и “Ёмкость аккумулятора” — 0.63, но мы склонны полагать,

что это случайное совпадение, связанное с размером нашей выборки (высокая корреляция между максимальным ISO и максимальным расширенным диапазоном ISO не является сильной, судя по рассчитанному VIF).

На основании визуального анализа целевой переменной, а также рассчитанного положительного коэффициента скошенности, распределение целевой переменной “Price” немного скошено вправо (асимметрично вправо), а значение эксцесса (Kurtosis) близко к 5.5, что указывает на то, что распределение данных далеко от нормального.

Что касается числовых переменных, то в их случае нельзя точно идентифицировать распределение, а коэффициенты асимметрии и эксцесса также достаточно велики. Из интересного: признаки “Видеозапись (качество)” и “Видеозапись (fps)” распределены менее равномерно, чем другие, и имеют ярко выраженные моды — 2160 пикселей и 30 fps.

Наконец, мы построили графики “ящики с усами” для числовых переменных (включая целевую) и выявили значительное количество выбросов в переменных “Price” и “Максимальный расширенный диапазон ISO”. Нами было выявлено, что одна камера Leica выпускается в ограниченном тираже и сильно дороже остальных, а также три камеры Nikon являются выбросами по “Максимальный расширенный диапазон ISO”. Далее весь анализ проводится без их учёта.

4 Выбор модели

Перед выбором модели необходимо проверить ключевые аспекты: функциональную форму, набор переменных, предпосылки Гаусса-Маркова, гипотезы о нормальности ошибок и значимости коэффициентов. Рассмотрены три спецификации модели (в полупологарифмической и лог-лог модели мы сначала использовали логарифмы, а затем стандартизировали регрессоры):

- Линейная,
- Полупологарифмическая (логарифм от зависимой переменной/логарифм от объясняемых переменных),
- Лог-лог (логарифмы от обеих сторон).

В анализ включена: 1) проверка нормальности остатков; 2) проверка мультиколлинеарности между признаками; 3) диагностика гетероскедастичности и проверка математического ожидания ошибок; 4) тест Рамсея для проверки корректности спецификации.

4.1 Тесты на нормальность случайных ошибок

Для проверки нормальности случайных ошибок мы использовали визуальный анализ (QQ-графики) и два теста: Харке-Бера и Колмогорова-Смирнова.

$$H_0 : X_i \sim N(\mu, \sigma^2), \quad i = 1, 2, \dots, n$$

где X_i — остатки модели, $N(\mu, \sigma^2)$ — нормальное распределение с математическим ожиданием μ и дисперсией σ^2 .

1. Тест Харке-Бера:

$$JB = \frac{n}{6} \left(sk^2 + \frac{1}{4}(k-3)^2 \right) \sim \chi^2$$

где n — количество наблюдений, sk — коэффициент асимметрии, k — коэффициент эксцесса.

2. Тест Колмогорова-Смирнова:

$$D_n = \sup_x |F_n(x) - F(x)|$$

где $F_n(x)$ — эмпирическая функция распределения, $F(x)$ — функция кумулятивного распределения для проверки отклонений от нормального распределения.

Результаты тестов на нормальность показывают, что для всех моделей p -значения тестов Jarque-Bera и Колмогорова-Смирнова настолько малы, что нулевая гипотеза о нормальности ошибок отклоняется. QQ-графики показывают, что линейная модель и модель с логарифмированными признаками X имеют значительные отклонения от нормальности, особенно в хвостах. Модель с логарифмированными признаками и таргетом даёт наилучший результат, с минимальными отклонениями. Все модели показывают нарушение нормальности ошибок, что снижает надёжность результатов для построения доверительных интервалов и оценки значимости коэффициентов.

4.2 Мультиколлинеарность

Из корреляционной матрицы и расчёта VIF видно, что для переменных, таких как Максимальный ISO ($VIF = 7.80$), Время работы батареи ($VIF = 4.17$) и Видео (качество) 2160 ($VIF = 3.22$), значения VIF превышают 1, что указывает на возможную мультиколлинеарность. Это может исказить оценки коэффициентов и снижать надёжность модели.

Так как VIF выше 10 обычно сигнализирует о сильной мультиколлинеарности, можно заключить, что в нашей модели она не представляет угрозы для точности

оценок.

4.3 Проверка на гетероскедастичность

После тестов на нормальность остатков необходимо проверить гетероскедастичность. Для этого используем визуальный анализ и статистические тесты: Бреуша-Пагана, Уайта, Вилкоксона и бутстреп-тест.

Тест Бройша-Пагана Для оценки гетероскедастичности в модели применяется тест Бройша-Пагана. Нулевая гипотеза в этом случае утверждает, что дисперсия ошибок не зависит от значений предсказанных переменных, то есть остатков.

Тестовая статистика рассчитывается по формуле:

$$BP = \frac{ESS}{2} \sim \chi_p^2$$

где ESS — это объясненная сумма квадратов, которая рассчитывается для вспомогательной регрессии ошибок:

$$\hat{\varepsilon}^2 = \gamma_0 + \gamma_1 Z_1 + \dots + \gamma_p Z_p + u$$

В которой $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$ — это оценка дисперсии ошибок, полученная из начальной регрессии.

Тест Уайта Тест Уайта используется для оценки гетероскедастичности и может быть полезен, когда модель имеет более сложную структуру ошибок. Тестовая статистика для теста Уайта выглядит следующим образом:

$$W = nR^2 \sim \chi_{m-1}^2$$

где R^2 рассчитывается для регрессии квадратов остатков (из начальной модели) на все регрессоры, их квадраты, попарные произведения и константу, а m — количество коэффициентов в этой регрессии.

Тест Вилкоксона Тест Вилкоксона используется для проверки медианных разностей между парами данных и является непараметрическим, что делает его полезным в случае, если распределение остатков не является нормальным. Нулевая гипотеза

этого теста заключается в том, что медиана разности между парами равна нулю:

H_0 : медиана разности = 0

Тестовая статистика для Вилкоксона рассчитывается как:

$$T = \sum_{i=1}^n \text{sign}(r_i) \cdot \text{rank}(r_i)$$

где r_i — разности между парами остатков, а $\text{rank}(r_i)$ — их ранги. Если результат теста указывает на отклонение от нуля, это может свидетельствовать о наличии гетероскедастичности.

Бутстреп-тест Бутстреп-тест основан на выборке с возвращением и позволяет оценить распределение статистики, используя множество повторных выборок из исходных данных.

Нулевая гипотеза для бутстреп-теста:

$$H_0 : \mu = 0$$

Тестовая статистика для бутстрапа рассчитывается следующим образом:

$$\hat{\mu}_{bootstrap} = \frac{1}{n} \sum_{i=1}^n x_i$$

где x_i — элементы бутстреп-выборки, а n — размер выборки. Этот метод особенно полезен в случаях, когда данные не соответствуют стандартным предположениям о нормальности и гетероскедастичности.

Тест Вилкоксона, использующийся для данных, не подчиняющихся нормальному распределению, показал, что медиана разности остатков не отклоняется от нуля, что подтверждает отсутствие значимых отклонений от нормальности, особенно для логарифмированных моделей. Бутстреп-тест, проводившийся с целью проверки устойчивости среднего значения остатков для модели, также показал p -значения больше 0.05, что подтверждает, что среднее значение для всех моделей не отличается от нуля. Графики остатков против прогнозов для линейной модели показали, что остатки сильно разрознены и не выглядят случайными, что может указывать на гетероскедастичность. В то же время, для логарифмированной модели остатки распределены более равномерно, что указывает на улучшение модели после логарифмирования данных и подтверждает, что логарифмирование может уменьшить гетероскедастичность.

4.4 Тест Рамсея на правильную спецификацию модели (для логарифмированной модели)

Тест Рамсея используется для проверки правильности спецификации модели, чтобы выяснить, не пропущены ли важные переменные или их степени, которые могут улучшить модель:

H_0 : спецификация модели правильная против H_1 : спецификация модели неправильная

Процесс теста включает: 1) оценку коэффициентов исходной модели, 2) сохранение столбца оцененных значений \hat{y} , 3) оценку коэффициентов для дополнительной регрессии (например, добавление квадрата \hat{y}).

Гипотеза проверяется через:

$$H_0 : \alpha_2 = 0$$

где α_2 — коэффициент при переменной \hat{y}^2 .

Тестовая статистика:

$$F = \frac{(RSS_R - RSS_{UR})/(m - 1)}{RSS_{UR}/(n - (k + m))}$$

где RSS_R и RSS_{UR} — остаточные суммы квадратов для исходной и расширенной моделей соответственно, m — количество добавленных переменных, n — количество наблюдений, k — количество коэффициентов в исходной модели.

В нашем случае тест Рамсея был пройден успешно ($pvalue = 0.08$) только при добавлении квадрата предсказанной оценки, а также квадратов регрессоров, что сильно усложняет интерпретируемость модели. **Мы решили остановиться на модели, где мы предсказываем логарифм цены, используя дамми-переменные и стандартизированные логарифмированные числовые переменные.**

4.5 Тест на адекватность модели

Для проверки адекватности модели используются основные и альтернативные гипотезы, которые выглядят следующим образом:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_i \neq 0$$

где $\beta_1, \beta_2, \dots, \beta_k$ — коэффициенты при независимых переменных. То есть, основная

гипотеза заключается в том, что все коэффициенты, кроме свободного, равны нулю. Поскольку при оценке модели с помощью метода наименьших квадратов (МНК) мы использовали встроенные в Python методы, значения тестовой статистики и p -значения уже есть в полученных результатах. Получившееся p -значение близко к нулю, что означает, что основная гипотеза отвергается на любом разумном уровне значимости (F-статистика = 15+). Таким образом, модель является статистически значимой.

4.3 Итоговая модель

После анализа мультиколлинеарности, гетероскедастичности и других факторов, мы пришли к решению использовать логарифмированную модель. Это было обусловлено тем, что линейная модель показала проблемы с мультиколлинеарностью, а также с неравномерностью дисперсии ошибок (гетероскедастичностью). В нашей модели категориальные переменные были представлены дамми-переменными, а числовые — стандартизированным логарифмом признака. Такая форма модели позволила улучшить предсказания, снизив эти проблемы и обеспечив стабильность модели.

$$\log(\text{Цена}) = 10.8499 + 0.3758 \cdot Z_{\text{Емкость аккумулятора}} + 0.1543 \cdot Z_{\text{Вес (в кг)}} + 1.2817 \cdot D_{\text{Leica}} + 0.6660 \cdot D_{2160p} + 1.0587 \cdot D_{3500p} + 1.0856 \cdot D_{4160p} + 1.0994 \cdot D_{4320p} - 0.6863 \cdot D_{720p} + 0.3359 \cdot D_{\text{Sony}} + \epsilon$$

z_{\dots} — стандартизированные логарифмы количественных переменных: сначала применяется $\log(1 + x)$, затем масштабирование через StandardScaler.

D_{\dots} — дамми-переменные для категориальных признаков; принимают значение 1, если условие выполняется, и 0 — иначе.

4.4 Предсказание модели

Переменная	Значение
Емкость аккумулятора	1900
Вес (в кг)	0.72
Производитель	Leica
Видео (качество)	2160

Таблица 1: Предложенные характеристики для нового фотоаппарата

Таким образом, согласно нашей модели, предсказанная цена фотоаппарата с такими характеристиками составляет 237776.99 рублей, а 95% доверительный интервал: от 94155.25 до 600475.18 рублей.

4.5 Интерпретация модели

Переменная	Коэфф.	Интерпретация
Емкость аккумуля.	0.3758	Рост log емкости на 1 std увеличивает цену на $\approx 45.6\%$
Вес (в кг)	0.1543	Рост log веса на 1 std увеличивает цену на $\approx 16.7\%$
Производитель Leica	1.2817	Leica дороже других на $\approx 260\%$
Производитель Sony	0.3359	Sony дороже базовых брендов на $\approx 39.9\%$
Видео = 2160p (4K)	0.6660	4K-видео увеличивает цену на $\approx 94.6\%$
Видео = 3500p	1.0587	Увеличивает цену на $\approx 188.3\%$
Видео = 4160p	1.0856	Увеличивает цену на $\approx 196.2\%$
Видео = 4320p (8K)	1.0994	Увеличивает цену на $\approx 200.3\%$
Видео = 720p	-0.6863	Снижает цену на $\approx 49.8\%$ по сравнению с 1080p

5 Квантильная регрессия

5.1 Почему выбрали квантильную регрессию?

В ходе первичного анализа данных мы обнаружили "выбросы" в целевой переменной, однако они скорее связаны с различиями в ценовых сегментах фотокамер, а не с аномальными значениями. В связи с этим для правильного анализа необходимо использовать квантильную регрессию, поскольку этот метод эффективно работает с выбросами и позволяет распределять значения по различным квантилям.

5.2 Модели

Результаты квартальных регрессий показывают, что коэффициенты значимо различаются по квантилям, что указывает на различия в влиянии факторов на цену в зависимости от ценового сегмента. Например, **Максимальный ISO** и **видеозапись в высоком качестве** оказывают большее влияние на цену в дорогих моделях (80-90% квантиль). **Производители** как **Sony** и **Nikon** увеличивают цену, особенно в высоких ценовых сегментах.

Некоторые характеристики, такие как **Размер экрана** и **Мегапиксели**, оказывают меньшее влияние на цену в более дешевых моделях. **Видео качества** и **время работы батареи** важны в бюджетных сегментах, тогда как в дорогих моделях критичны такие параметры и более высокое качество съемки.

Анализ показывает, что с увеличением цены в модели становится важным качество видеосъемки и характеристики, связанные с производителем, что подчеркивает зна-

чимость технологий для более дорогих сегментов. Квантильная регрессия помогает выявить, как различные характеристики влияют на цену в зависимости от сегмента рынка.

Вот пример одного уравнения регрессии:

90 %:

$$\begin{aligned} \log(\text{Price}) = & 12.0664 + 0.1000 \cdot z_{ISO} - 0.0516 \cdot z_{ISOFull} + 0.2180 \cdot z_{Shots} + 0.3437 \cdot \\ & z_{Accumulator} + 0.1701 \cdot z_{Weight} + 0.0219 \cdot z_{Screen} - 0.0860 \cdot z_{Megapixels} + 0.0876 \cdot z_{FPS} + 0.1908 \cdot \\ & D_{Fujifilm} + 1.2142 \cdot D_{Leica} + 0.3647 \cdot D_{Nikon} + 0.2920 \cdot D_{Panasonic} + 0.8212 \cdot D_{Sony} - 0.2216 \cdot \\ & D_{protection.} + 0.2504 \cdot D_{FPS24} + 0.2577 \cdot D_{FPS30} + 0.2944 \cdot D_{FPS60} + 0.9190 \cdot D_{Video3500p} + \\ & 1.5813 \cdot D_{Video4160p} + 1.6048 \cdot D_{Video4320p} - 0.2790 \cdot D_{Video720p} + \varepsilon \end{aligned}$$

5.3 Визуальный анализ

После оценки квантильной регрессии выделяются ключевые моменты. Например, коэффициент для "Максимальный ISO" (0.1149, p-value = 0.000) увеличивается с ростом квантиля, что подтверждает его влияние на цену в верхних квантилях. Напротив, "Максимальный расширенный диапазон ISO" имеет отрицательное влияние, особенно на 90% квантиле (-0.0516), что снижает цену.

Переменные, связанные с производителями, например, Fujifilm (1.2142 на 90% квантиле), показывают положительное влияние на цену, а переменные, как "Видеоозапись fps_24" (-1.0762) и "fps_30" (-0.6595), уменьшают стоимость моделей.

Эти результаты показывают, как квантильная регрессия позволяет выявить изменения зависимостей для разных уровней цен, предоставляя полезную информацию для анализа рынка.

6 Вывод

Таким образом, в рамках работы была успешно оценена модель ценообразования для фотокамер с использованием методов множественной регрессии. На начальном этапе был произведен сбор данных и их подробный количественный и визуальный анализ, а также выполнен отбор независимых переменных.

Для построения модели были проверены основные предпосылки для применения метода наименьших квадратов, включая мультиколлинеарность, гетероскедастичность и другие важные факторы. Далее был выбран подходящий тип модели, что позволило выявить статистически значимые переменные и построить предсказания для цен на фотокамеры с различными характеристиками.

В качестве дополнительного шага, была использована квантильная регрессия, что позволило более глубоко проанализировать влияние отдельных факторов на цену камеры в зависимости от категории. Данная модель продемонстрировала важность таких характеристик, как разрешение видео и емкость аккумулятора, для формирования цен в различных сегментах рынка.

Приложение (основные модели и тесты, подробнее в iрunb - файле)

Результаты линейной модели: OLS Regression Results

Dep. Variable:	Цена	R-squared:	0.782
Model:	OLS	Adj. R-squared:	0.719
Method:	Least Squares	F-statistic:	12.41
Date:	Thu, 08 May 2025	Prob (F-statistic):	4.18e-17
Time:	18:43:55	Log-Likelihood:	-1229.8
No. Observations:	99	AIC:	2526.
Df Residuals:	76	BIC:	2585.
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	1.633e+05	4.74e+04	3.443	0.001	6.88e+04	2.58e+05
Максимальный ISO	167.4900	1.69e+04	0.010	0.992	-3.35e+04	3.39e+04
Максимальный расширенный диапазон ISO	2.751e+04	1.89e+04	1.457	0.149	-1.81e+04	6.51e+04
Время работы батареи (shots)	1164.3941	8942.268	0.130	0.897	-1.66e+04	1.9e+04
Емкость аккумулятора	1.914e+04	1.51e+04	1.263	0.210	-1.1e+04	4.93e+04
Вес (в kg)	2.486e+04	1.18e+04	2.105	0.039	1336.849	4.84e+04
Размер экрана (без ")	1.181e+04	1.04e+04	1.138	0.259	-8861.250	3.25e+04
Мегапиксели (основная камера)	2.536e+04	1.23e+04	2.056	0.043	794.968	4.99e+04
Непрерывная съемка FPS	2.002e+04	8874.698	2.358	0.021	3248.750	3.86e+04
Производитель_Fujifilm	-2575.2610	3.62e+04	-0.071	0.943	-7.46e+04	6.95e+04
Производитель_Leica	2.314e+05	3.46e+04	6.687	0.000	1.62e+05	3e+05
Производитель_Nikon	-4883.0063	2.69e+04	-0.181	0.857	-5.06e+04	4.88e+04
Производитель_Panasonic	-1216.5831	3.03e+04	-0.040	0.968	-6.13e+04	5.89e+04
Производитель_Sony	3.063e+04	2.79e+04	1.097	0.276	-2.5e+04	8.62e+04
Наличие пыле/влаго защиты_1	-1.384e+04	2.39e+04	-0.580	0.564	-6.14e+04	3.37e+04
Видеозапись fps_24	-6.148e+04	5.46e+04	-1.125	0.264	-1.7e+05	4.73e+04
Видеозапись fps_30	-5.596e+04	4.33e+04	-1.292	0.200	-1.42e+05	3.83e+04
Видеозапись fps_60	-5.23e+04	4.12e+04	-1.268	0.209	-1.34e+05	2.98e+04
Видео (качество)_2160	1.681e+04	2.46e+04	0.684	0.496	-3.22e+04	6.58e+04
Видео (качество)_3500	8.114e+04	4.53e+04	1.792	0.077	-9025.488	1.71e+05
Видео (качество)_4160	5.83e+04	6e+04	0.971	0.335	-6.13e+04	1.78e+05
Видео (качество)_4320	1.311e+05	4.43e+04	2.959	0.004	4.28e+04	2.19e+05
Видео (качество)_720	3.037e+04	7.21e+04	0.421	0.675	-1.13e+05	1.74e+05

Omniбус: 39.122 Durbin-Watson: 1.749
Prob(Omniбус): 0.000 Jarque-Bera (JB): 191.294
Skew: 1.122 Prob(JB): 2.89e-42
Kurtosis: 9.430 Cond. No. 20.5

Результаты модели с логарифмическими признаками X: OLS Regression Results

Dep. Variable:	Цена	R-squared:	0.751
Model:	OLS	Adj. R-squared:	0.678
Method:	Least Squares	F-statistic:	10.40
Date:	Thu, 08 May 2025	Prob (F-statistic):	4.08e-15
Time:	18:43:55	Log-Likelihood:	-1246.5
No. Observations:	99	AIC:	2530.
Df Residuals:	76	BIC:	2599.
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	1.661e+05	4.95e+04	3.355	0.001	6.75e+04	2.65e+05
Максимальный ISO	-1.185e+04	2.28e+04	-0.520	0.604	-5.72e+04	3.35e+04
Максимальный расширенный диапазон ISO	2.093e+04	2.24e+04	0.935	0.353	-2.37e+04	6.55e+04
Время работы батареи (shots)	4100.3041	1.09e+04	0.377	0.708	-1.76e+04	2.58e+04
Емкость аккумулятора	1.567e+04	1.66e+04	0.941	0.350	-1.75e+04	4.88e+04
Вес (в kg)	2.544e+04	1.31e+04	1.940	0.056	-678.797	5.16e+04
Размер экрана (без ")	1.293e+04	1.1e+04	1.172	0.245	-9042.503	3.49e+04
Мегапиксели (основная камера)	1.009e+04	1.21e+04	0.834	0.407	-1.4e+04	3.42e+04
Непрерывная съемка FPS	1.81e+04	1.84e+04	0.969	0.335	-1.07e+04	3.89e+04
Производитель_Fujifilm	-2.638e+04	4.19e+04	-0.630	0.531	-1.1e+05	5.7e+04
Производитель_Leica	2.289e+05	3.83e+04	5.969	0.000	1.53e+05	3.05e+05
Производитель_Nikon	-1.821e+04	2.92e+04	-0.630	0.727	-6.83e+04	4.79e+04
Производитель_Panasonic	-1.789e+04	3.35e+04	-0.537	0.593	-8.46e+04	4.87e+04
Производитель_Sony	4.468e+04	2.98e+04	1.499	0.138	-1.47e+04	1.04e+05
Наличие пыле/влаго защиты_1	-5818.8810	2.55e+04	-0.228	0.820	-5.66e+04	4.49e+04
Видеозапись fps_24	-1.854e+05	5.67e+04	-1.658	0.067	-2.18e+05	7599.690
Видеозапись fps_30	-8.74e+04	4.44e+04	-1.968	0.053	-1.76e+05	1054.392
Видеозапись fps_60	-7.229e+04	4.27e+04	-1.692	0.095	-1.57e+05	1.28e+04
Видео (качество)_2160	4.054e+04	2.94e+04	1.378	0.172	-1.81e+04	9.91e+04
Видео (качество)_3500	1.323e+05	4.71e+04	2.800	0.006	3.85e+04	2.26e+05
Видео (качество)_4160	9.546e+04	6.55e+04	1.457	0.149	-3.5e+04	7.26e+05
Видео (качество)_4320	1.824e+05	4.94e+04	3.694	0.000	8.41e+04	2.81e+05
Видео (качество)_720	4.734e+04	7.81e+04	0.606	0.546	-1.08e+05	2.03e+05

Omniбус: 21.189 Durbin-Watson: 1.725
Prob(Omniбус): 0.000 Jarque-Bera (JB): 53.632
Skew: 0.711 Prob(JB): 2.26e-12
Kurtosis: 6.313 Cond. No. 21.4

Результаты модели с логарифмированным таргетом: OLS Regression Results

Dep. Variable:	Цена	R-squared:	0.829
Model:	OLS	Adj. R-squared:	0.788
Method:	Least Squares	F-statistic:	16.75
Date:	Thu, 08 May 2025	Prob (F-statistic):	7.54e-21
Time:	18:43:55	Log-Likelihood:	-46.032
No. Observations:	99	AIC:	130.1
Df Residuals:	76	BIC:	197.8
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	11.2592	0.275	40.949	0.000	10.712	11.807
Максимальный ISO	-0.0056	0.008	-0.057	0.954	-0.201	0.190
Максимальный расширенный диапазон ISO	0.1320	0.109	1.206	0.232	-0.086	0.350
Время работы батареи (shots)	0.0657	0.052	1.266	0.209	-0.038	0.169
Емкость аккумулятора	0.2263	0.088	2.577	0.012	0.051	0.401
Вес (в kg)	0.2019	0.068	2.948	0.004	0.065	0.338
Размер экрана (без ")	0.0057	0.060	0.095	0.925	-0.114	0.126
Мегапиксели (основная камера)	0.1375	0.072	1.922	0.058	-0.005	0.280
Непрерывная съемка FPS	0.0571	0.051	1.109	0.271	-0.045	0.168
Производитель_Fujifilm	0.2037	0.210	0.971	0.334	-0.214	0.621
Производитель_Leica	1.2605	0.201	6.323	0.000	0.869	1.668
Производитель_Nikon	0.1083	0.156	0.642	0.523	-0.211	0.412
Производитель_Panasonic	-0.0089	0.175	-0.051	0.960	-0.357	0.340
Производитель_Sony	0.2267	0.162	1.400	0.165	-0.096	0.549
Наличие пыле/влаго защиты_1	-0.1377	0.138	-0.995	0.323	-0.414	0.138
Видеозапись fps_24	-0.3500	0.317	-1.105	0.273	-0.981	0.281
Видеозапись fps_30	-0.3561	0.251	-1.418	0.160	-0.856	0.144
Видеозапись fps_60	-0.2515	0.239	-1.052	0.296	-0.728	0.225
Видео (качество)_2160	0.6290	0.143	4.413	0.000	0.345	0.913
Видео (качество)_3500	0.7905	0.262	3.012	0.004	0.268	1.313
Видео (качество)_4160	0.8892	0.348	2.554	0.013	0.196	1.583
Видео (качество)_4320	0.9974	0.257	3.884	0.000	0.486	1.509
Видео (качество)_720	-0.5969	0.418	-1.427	0.158	-1.430	0.236

Omniбус: 27.700 Durbin-Watson: 2.161
Prob(Omniбус): 0.000 Jarque-Bera (JB): 53.558
Skew: -1.104 Prob(JB): 2.34e-12
Kurtosis: 5.848 Cond. No. 20.5

Результаты модели с логарифмированными признаками и таргетом: OLS Regression Results

Dep. Variable:	Цена	R-squared:	0.838
Model:	OLS	Adj. R-squared:	0.792
Method:	Least Squares	F-statistic:	17.93
Date:	Thu, 08 May 2025	Prob (F-statistic):	9.68e-22
Time:	18:43:55	Log-Likelihood:	-43.216
No. Observations:	99	AIC:	132.4
Df Residuals:	76	BIC:	192.1
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	11.2355	0.261	43.096	0.000	10.716	11.755
Максимальный ISO	-0.0082	0.120	-0.068	0.946	-0.247	0.231
Максимальный расширенный диапазон ISO	0.1218	0.118	1.033	0.305	-0.113	0.357
Время работы батареи (shots)	0.0836	0.057	1.458	0.149	-0.031	0.198
Емкость аккумулятора	0.2007	0.088	2.358	0.021	0.032	0.381
Вес (в kg)	0.2099	0.069	3.039	0.003	0.072	0.347
Размер экрана (без ")	0.0254	0.058	0.438	0.663	-0.090	0.141
Мегапиксели (основная камера)	0.0759	0.064	1.192	0.237	-0.051	0.203
Непрерывная съемка FPS	0.0066	0.055	0.120	0.905	-0.103	0.116
Производитель_Fujifilm	0.1714	0.221	0.777	0.440	-0.268	0.611
Производитель_Leica	1.2788	0.202	6.332	0.000	0.877	1.681
Производитель_Nikon	0.0956	0.154	0.622	0.536	-0.210	0.402
Производитель_Panasonic	0.0314	0.176	0.178	0.859	-0.320	0.382
Производитель_Sony	0.3428	0.157	2.184	0.032	0.030	0.655
Наличие пыле/влаго защиты_1	-0.1612	0.134	-1.201	0.233	-0.429	0.106
Видеозапись fps_24	-0.4462	0.299	-1.494	0.139	-1.041	0.149
Видеозапись fps_30	-0.4095	0.234	-1.751	0.084	-0.875	0.056
Видеозапись fps_60	-0.3027	0.225	-1.346	0.182	-0.751	0.145
Видео (качество)_2160	0.6800	0.155	4.454	0.000	0.381	0.908
Видео (качество)_3500	0.9718	0.248	3.915	0.000	0.477	1.466
Видео (качество)_4160	0.9626	0.345	2.790	0.007	0.275	1.650
Видео (качество)_4320	1.1534	0.260	4.434	0.000	0.635	1.671
Видео (качество)_720	-0.4810	0.411	-1.170	0.246	-1.300	0.338

Omniбус: 22.730 Durbin-Watson: 2.100
Prob(Omniбус): 0.000 Jarque-Bera (JB): 36.916
Skew: -0.986 Prob(JB): 9.63e-09
Kurtosis: 5.249 Cond. No. 21.4

Результаты теста на нормальность (Jarque-Bera и Kolmogorov-Smirnov):

Линейная модель:

Jarque-Bera p-value: 2.8904702881056145e-42

Kolmogorov-Smirnov p-value: 6.283238234818095e-25

Модель с логарифмированными признаками X:

Jarque-Bera p-value: 2.259353213367746e-12

Kolmogorov-Smirnov p-value: 6.283238234818095e-25

Модель с логарифмированным таргетом:

Jarque-Bera p-value: 2.3443816151461353e-12

Kolmogorov-Smirnov p-value: 8.442635114436493e-06

Модель с логарифмированными признаками и таргетом:

Jarque-Bera p-value: 9.631553205753028e-09

Kolmogorov-Smirnov p-value: 2.578696027373478e-06

RESET-тест Рамсея с робастными ошибками (HC3) для лог-лог модели:
F-статистика: 1.7733

p-value: 0.089846

/usr/local/lib/python3.11/dist-packages/statsmodels/regression/lin
warnings.warn('F test for comparison is likely invalid with ' +

Wilcoxon тест для линейной модели:

P-value: 0.71661

95% доверительный интервал (percentile CI): (-106401.22197, 123619.93726)

Wilcoxon тест для модели с логарифмированными признаками X:

P-value: 0.84504

95% доверительный интервал (percentile CI): (-119869.42234, 155270.57751)

Wilcoxon тест для лог-лог модели:

P-value: 0.50277

95% доверительный интервал (percentile CI): (-0.78607, 0.59391)

Результаты для модели Линейная модель:

Breusch-Pagan p-value: 0.04747767429383432

White's test p-value: 0.4527212632423215

Результаты для модели Логарифмированная модель:

Breusch-Pagan p-value: 0.016185714381707288

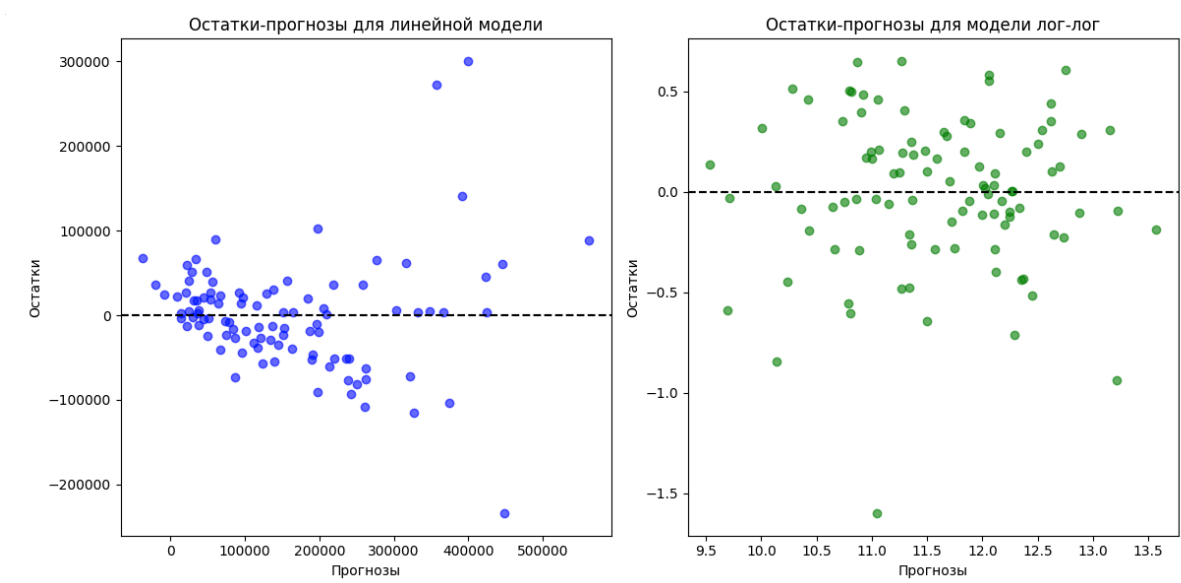
White's test p-value: 0.4527212632423215

Результаты для модели Лог-лог:

Breusch-Pagan p-value: 0.5933005308419648

White's test p-value: 0.4527212632423215

Приложение (основные модели и тесты, подробнее в iрупb - файле)



OLS Regression Results					
Dep. Variable:	Цена	R-squared:	0.806		
Model:	OLS	Adj. R-squared:	0.786		
Method:	Least Squares	F-statistic:	41.10		
Date:	Thu, 08 May 2025	Prob (F-statistic):	5.74e-28		
Time:	18:44:01	Log-Likelihood:	-52.265		
No. Observations:	99	AIC:	124.5		
Df Residuals:	89	BIC:	150.5		
Df Model:	9				
Covariance Type: nonrobust					
	coef	std err	t	P> t	[0.025 0.975]
Емкость аккумулятора	0.3758	0.060	6.290	0.000	0.257 0.495
Вес (в kg)	0.1543	0.059	2.634	0.010	0.038 0.271
Производитель_Leica	1.2817	0.158	8.135	0.000	0.969 1.595
Видео (качество)_2160	0.6660	0.110	6.054	0.000	0.447 0.885
Видео (качество)_3500	1.0587	0.209	5.054	0.000	0.642 1.475
Видео (качество)_4160	1.0856	0.272	3.990	0.000	0.545 1.626
Видео (качество)_4320	1.0994	0.177	6.214	0.000	0.748 1.451
Видео (качество)_720	-0.6863	0.319	-2.149	0.034	-1.321 -0.052
Производитель_Sony	0.3359	0.116	2.896	0.005	0.105 0.566
const	10.8499	0.094	114.991	0.000	10.662 11.037
Omnibus:	9.343	Durbin-Watson:	2.085		
Prob(Omnibus):	0.009	Jarque-Bera (JB):	9.672		
Skew:	-0.600	Prob(JB):	0.00794		
Kurtosis:	3.951	Cond. No.	9.72		

Предсказанная цена: 237776.9974680443 ₺
95% доверительный интервал: от 94155.26565878604 ₺ до 600475.1845723516 ₺