

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ



ЭКОНОМЕТРИКА-2  
ПРОЕКТНАЯ РАБОТА

---

Оценивание гедонистической  
ценовой функции сумок

---

Студент	Вклад	Обязанности
Красавцева Дарья, ЭАД	33,(3)%	Реализация и интерпретация модели
Федорова Анастасия, ЭАД	33,(3)%	Подготовка данных, реализация квантильной регрессии
Шмелева Анна, ЭАД	33,(3)%	Написание текста отчёта, проверка данных

# Содержание

<b>1</b>	<b>Описание данных</b>	<b>2</b>
1.1	Товар . . . . .	2
1.2	Источник данных . . . . .	2
1.3	Переменные для анализа . . . . .	2
1.4	Описательные статистики . . . . .	2
1.5	Квантили . . . . .	3
1.6	Средние . . . . .	3
1.7	Box plots . . . . .	4
1.8	Гистограммы распределений переменных . . . . .	5
1.9	Матрица корреляции . . . . .	6
<b>2</b>	<b>Подбор функциональной формы</b>	<b>7</b>
2.1	Извлечение признаков . . . . .	7
2.2	Линейная модель . . . . .	7
2.3	Полулогарифмическая модель . . . . .	8
2.4	Линейная в логарифмах модель . . . . .	10
2.5	Сравнение моделей . . . . .	11
<b>3</b>	<b>Предпосылки теоремы Гаусса-Маркова</b>	<b>11</b>
3.1	Теорема . . . . .	11
3.2	Линейная зависимость по параметрам . . . . .	12
3.3	Линейная независимость признаков . . . . .	12
3.4	Несмещенность ошибок . . . . .	12
3.5	Гомоскедастичность . . . . .	13
3.6	Некоррелируемость ошибок . . . . .	14
<b>4</b>	<b>Гипотеза о нормальности случайной ошибки</b>	<b>15</b>
<b>5</b>	<b>Гипотеза о значимости коэффициентов и модели в целом</b>	<b>16</b>
5.1	Значимость коэффициентов . . . . .	16
5.2	Значимость модели . . . . .	17
<b>6</b>	<b>Интерпретация итоговой модели</b>	<b>17</b>
<b>7</b>	<b>Прогноз цены своего товара</b>	<b>19</b>
7.1	Предсказание на реальных данных . . . . .	19
7.2	Предсказание на смоделированных данных . . . . .	19
<b>8</b>	<b>Квантильная регрессия</b>	<b>20</b>
8.1	Анализ результатов квантильной регрессии для разных кван- тилей . . . . .	20
8.2	Квантильная регрессия с бутстрапированными выборками	22
8.3	Тест Вальда . . . . .	23
8.4	Графики для каждого предиката . . . . .	24

# 1 Описание данных

## 1.1 Товар

В качестве товара в данной работе рассматриваются сумки и чемоданы.

## 1.2 Источник данных

Данные были собраны с сайта [ЦУМ](#) — крупного российского ритейлера люксовых товаров.

Метод сбора: Парсинг веб-страниц с использованием Python (библиотеки requests, BeautifulSoup, pandas).

## 1.3 Переменные для анализа

Собранные данные включают следующие переменные:

Переменная	Тип данных	Описание	Единицы измерения
brand	Категориальная	Бренд сумки	Текст
color	Категориальная	Цвет	Текст
material	Категориальная	Состав материала	Текст
design_country	Категориальная	Страна дизайна	Текст
manufacture_country	Категориальная	Страна производства	Текст
article	Числовая (ID)	Артикул товара	Число
price	Числовая	Цена	Рубли
height	Числовая	Высота сумки	Сантиметры (см)
width	Числовая	Ширина сумки	Сантиметры (см)
depth	Числовая	Глубина сумки	Сантиметры (см)
included_items	Категориальная	Что входит в комплект	Текст
is_suitcase	Бинарная	1 если это чемодан	1 или 0

## 1.4 Описательные статистики

Таблица 1: Описательная статистика числовых переменных

	price	height	width	depth	is_suitcase
count	650.00	650.00	650.00	650.00	650.00
mean	200016.58	22.52	29.84	10.56	0.02
std	428898.52	9.87	9.41	5.18	0.12
min	9500.00	7.00	9.00	1.00	0.00
25%	44600.00	15.00	23.00	7.00	0.00
50%	72450.00	20.00	28.00	10.00	0.00
75%	173625.00	28.00	36.00	13.00	0.00
max	3215000.00	76.00	66.00	47.00	1.00

Анализ описательных статистик:

- price

Средняя цена сумок в данных — около 200 тысяч рублей, но это не слишком репрезентативно так как есть несколько супердорогих моделей

(до 3.2 миллиона), обычная средняя величина сильно завышена. более реальная картина видна по медиане — типичная сумка стоит 72 тысячи и половина товаров дешевле этой отметки. брендов

- размеры

Если брать типичную сумку по медиане, то это: Высота 20 см , Ширина 28 см , Глубина 10 см

Но в данных есть пара особенностей:

1. Встречаются гиганты высотой 76 см и глубиной 47 см — скорее всего это чемоданы или дорожные сумки

2. Есть и "плоские" варианты глубиной всего 1 см (может быть речь о клатчах)

- чемоданы

Судя по данным, чемоданов совсем мало. Всего 2

В итоге можно сказать что цены сильно перекошены из-за пары сверхдорогих позиций, размеры в основном стандартные, но есть несколько аномалий

## 1.5 Квантили

Квартильные значения:

	price	height	width	depth
0.25	44600.0	15.0	23.0	7.0
0.50	72450.0	20.0	28.0	10.0
0.75	173625.0	28.0	36.0	13.0

Межквартильный размах:

price	129025.0
height	13.0
width	13.0
depth	6.0

Анализ:

- price

Разница между бюджетными и премиальными сумками внутри среднего диапазона колоссальная ( $IQR = 129\,025$  огромный разброс), это подтверждает что цены распределены неравномерно: есть много дешёвых вариантов и резкий скачок к дорогим

- Габариты

Высота и ширина имеют одинаковый  $IQR$  (13 см), но разные  $Q1$  и  $Q3$ . Ширина в целом больше высоты.

Глубина варьируется слабее ( $IQR = 6$  см) а медиана — 10 см, что соответствует компактным повседневным моделям.

## 1.6 Средние

Анализ:

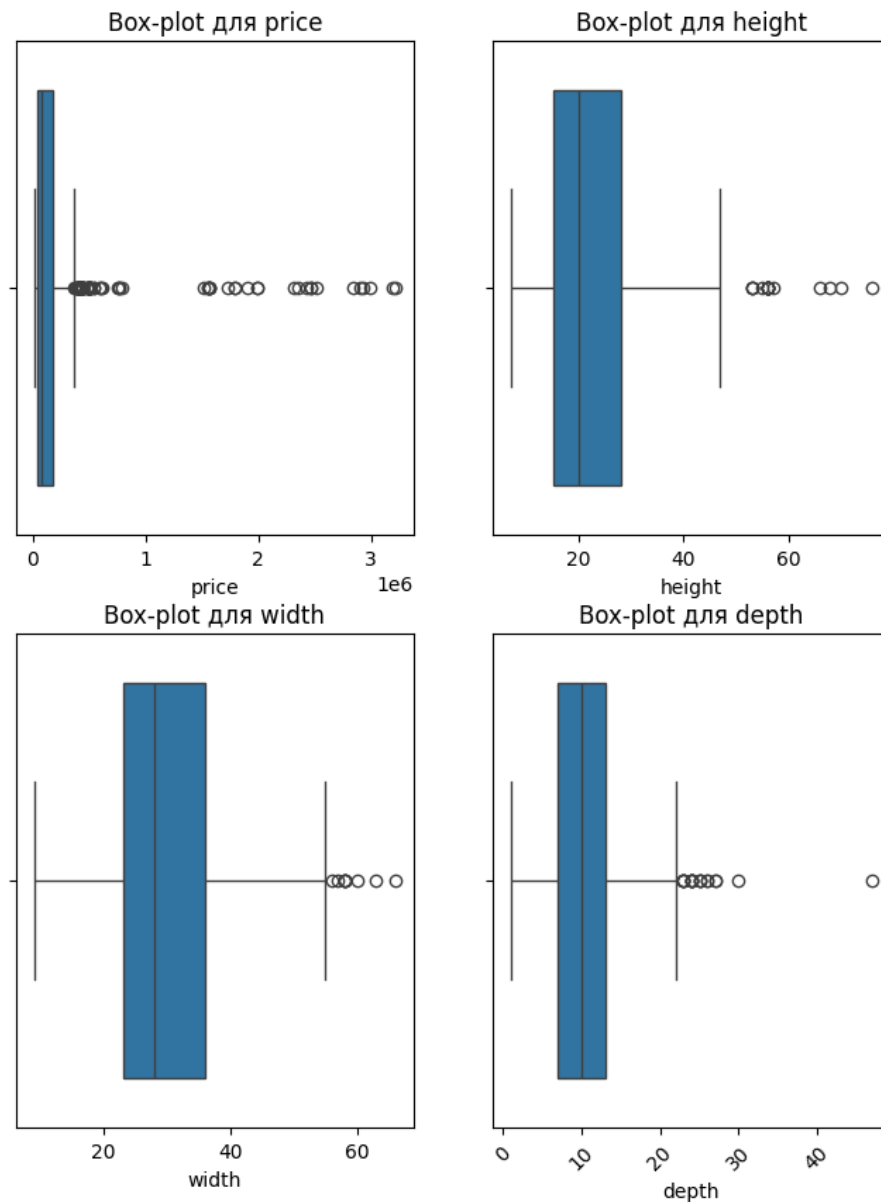
	Mean	Median	Mode
price	200016.576923	72450.0	99500.0
height	22.515385	20.0	19.0
width	29.835385	28.0	24.0
depth	10.557231	10.0	10.0

По таблице видно что у столбца depth симметричное распределение, то есть медиана мода и среднее очень близки друг к другу.

У столбцов height, width ассиметричное распределение с положительной ассиметрией. При положительной асимметрии среднее значение будет выше медианы, которая, в свою очередь, будет выше моды.

У столбца price ассиметричное распределение.

## 1.7 Box plots



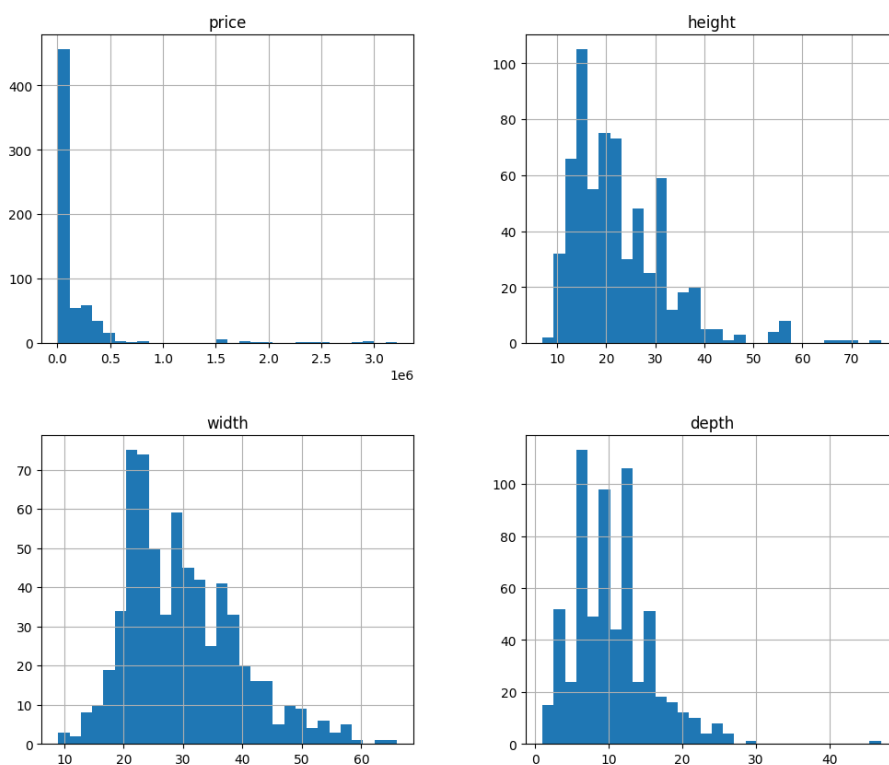
Посмотрим на выбросы  
Они есть во всех столбцах.

В столбце price: это вполне логично так как идет смешение люксового и более простого сегментов сумок в зависимости от бренда.

В столбцах height width depth тоже есть вбросы на графике, это закономерно так как возможно это выделяются чемоданы или особенно большие сумки. При этом выбросы есть именно с большей стороны, то есть слишком маленьких значений которые могли бы считаться выбросами нету тут.

## 1.8 Гистограммы распределений переменных

Гистограммы распределения переменных



Про асимметрию:

Если смотреть только на графики, то можно сказать что симметричное распределение есть у depth (если не считать один выброс справа) и width примерно. Гистограмма выглядит как колокол (нормальное распределение), где среднее, медиана и мода совпадают. Это указывает на отсутствие асимметрии и равномерное распределение данных.

У графиков price и height присутствует положительная асимметрия (на гистограмме наблюдается длинный правый хвост), что указывает на то, что большинство значений сосредоточено в нижней части диапазона, а более высокие значения встречаются реже.

Про островершинность:

Положительную островершинность, то есть более высокую и острую вершину с тяжелыми хвостами, мы можем наблюдать у price и height.

Возможно еще у depth. Это говорит о том, что данные имеют больше выбросов и экстремальных значений по сравнению с нормальным распределением.

Отрицательную островершинность, то есть гистограмма выглядит более плоской и широкой, чем нормальное распределение, можно наблюдать у width. Это говорит о том, что данные имеют меньше выбросов по сравнению с нормальным распределением.

Skewness:

price	4.783592
height	1.558678
width	0.806900
depth	1.153825

Kurtosis:

price	24.487452
height	3.838491
width	0.598793
depth	3.880968

Примерно то, что мы видели на графиках, мы видим и на числах.

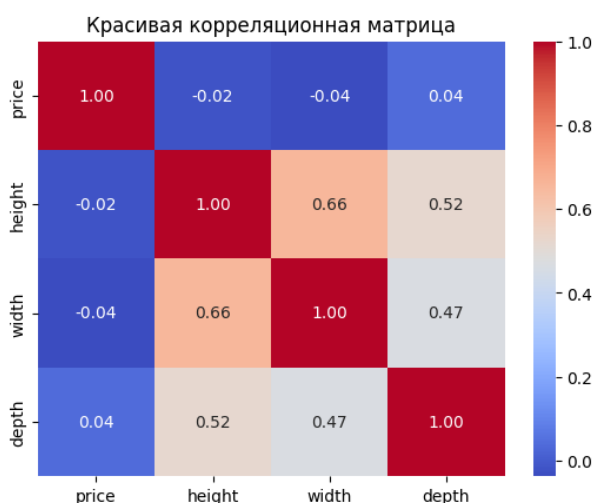
Выводы можно сделать следующие.

Если коэффициент асимметрии положителен, это может указывать на наличие большого числа низких значений в то время как высокие значения, выбросы, тянут среднее вверх. Если коэффициент отрицателен, это может говорить о том, что большинство значений сосредоточено в верхней части диапазона, а низкие значения и выбросы тянут среднее вниз.

Про островершинность: Положительное значение указывает на то, что данные имеют тенденцию к большему количеству выбросов. Отрицательное значение говорит о том, что данные более равномерно распределены и имеют меньше выбросов.

У нас все положительное !

## 1.9 Матрица корреляции



Общие результаты по матрице корреляций:

Сильная связь (больше 0.2) не наблюдается.

Средняя взаимосвязь (от 0.1 до 0.2) тоже не наблюдается.  
Низкая связь (меньше 0.1) - у цены со всеми переменными

## 2 Подбор функциональной формы

### 2.1 Извлечение признаков

В исходном датасете в столбце `included_items` в разном формате были перечислены аксессуары, прилагающиеся к сумке («второй ремешок длиной 41 см», «съёмная часть ремешка длиной 39 см»). Мы решили сделать бинарные признаки по ключевым словам («второй ремешок», «брелок» ...).

Дамми-признаки по аксессуарам: «keychain», «lock», «dust\_cover», «mirror», «second\_belt», «makeup\_bag».

Также в данных есть категориальные признаки: бренд, материал, страна дизайна, страна изготовления и цвет. `OneHotEncode`-им качественные признаки. У нас очень много уникальных значений материала. Безусловно, кожа страуса и кожа аллигатора влияют на цену, но, поскольку такие материалы встречаются редко, мы решили объединить их в одну «редкую» категорию. Аналогично поступим со всеми редкими категориями - будем считать категорию редкой, если в данных меньше двадцати сумок с такой категорией.

Чтобы избежать мультиколлинеарности, редкая категория не будет иметь отдельного дамми-признака, так как она выражается через константу и остальные дамми-признаки по этому категориальному признаку.

### 2.2 Линейная модель

Оценим линейную модель, объясняющую цену сумки её размерами и перечисленными дамми-признаками.

Уравнение оцененной модели представлено в Таблице 2

$$R^2 = 0.381$$

Получилось неплохое значение коэффициента детерминации. F-тест показывает, что на 5% уровне значимости наша модель в целом значима.

На этом уровне **значимы** следующие коэффициенты: `keychain`, `lock`, `mirror`, `second_belt`, `brand_Borbonese`, `brand_Bottega Veneta`, `brand_Furla`, `brand_MARC JACOBS (THE)`, `brand_MICHAEL Michael Kors`, `material_Кожа натуральная: 100%`, `material_Текстиль: 100%`, `design_country_США`, `manufacture_country_Вьетнам`, `manufacture_country_Италия`, `manufacture_country_Мадагаскар`.

Все цвета оказались незначимые, что логично: товар часто бывает представлен в разных цветах, отчего его цена не меняется. Оказались



Таблица 2: Уравнение линейной модели и p-value t-тестов.

	признак	значение	p-value
	height	-2385.06	3.13e-01
	width	-288.22	8.93e-01
	depth	6522.68	5.55e-02
	is_suitcase	69059.66	6.29e-01
	keychain	198100.96	3.06e-03
	lock	385254.46	1.87e-02
	dust_cover	110897.51	3.24e-01
	mirror	747632.71	3.48e-04
	second_belt	216525.77	9.28e-03
	makeup_bag	-30535.33	6.63e-01
	brand_Borbonese	-321208.47	1.14e-06
	brand_Bottega Veneta	299929.92	2.13e-05
	brand_Coccinelle	-11223.59	8.61e-01
	brand_Dolce & Gabbana	103037.16	1.46e-01
	brand_Furla	-207250.51	2.57e-03
	brand_MARC JACOBS (THE)	-498528.47	2.52e-04
	brand_MICHAEL Michael Kors	-769652.21	7.86e-11
	brand_Sans-Arcidet	138439.80	4.39e-01
	color_Бежевый	2055.99	9.66e-01
	color_Голубой	105572.68	2.03e-01
	color_Коричневый	-57447.86	3.55e-01
	color_Кремовый	-9327.14	8.79e-01
	color_Светло-бежевый	-39368.31	6.30e-01
	color_Черный	-19788.93	5.69e-01
	material_Кожа натуральная: 100%	-213314.69	1.29e-08
	material_Текстиль: 100%	-279275.46	5.98e-07
	design_country_Италия	62310.52	3.13e-01
	design_country_США	751684.48	2.67e-11
	design_country_Соединенное Королевство	1456.77	9.86e-01
	design_country_Франция	47399.05	4.80e-01
	manufacture_country_Вьетнам	-259512.86	1.68e-02
	manufacture_country_Италия	218130.22	7.47e-06
	manufacture_country_Камбоджа	-138358.56	1.94e-01
	manufacture_country_Китай	-43183.16	4.23e-01
	manufacture_country_Мадагаскар	-353426.57	3.70e-02
	manufacture_country_Тунис	-41997.46	6.45e-01
	const	80977.97	5.52e-01

значимы практически все крупные бренды, из распространенных стран дизайна оказалась значимой только США, из крупных стран производства значима только половина.

## 2.3 Полулогарифмическая модель

Оценим полулогарифмическую модель, объясняющую логарифм цены ( $\ln(price)$ ) теми же признаками.

Уравнение оцененной модели представлено в Таблице 3

Таблица 3: Уравнение полулогарифмической модели и p-value t-тестов.

признак	значение	p-value
height	-0.00	2.83e-01
width	0.00	5.46e-01
depth	0.02	3.65e-03
is_suitcase	0.19	4.31e-01
keychain	0.33	3.86e-03
lock	0.54	5.27e-02
dust_cover	0.25	1.97e-01
mirror	1.01	4.89e-03
second_belt	0.32	2.65e-02
makeup_bag	0.03	8.07e-01
brand_Borbonese	-0.90	4.68e-15
brand_Bottega Veneta	0.90	2.75e-13
brand_Coccinelle	-0.62	1.92e-08
brand_Dolce & Gabbana	0.58	1.87e-06
brand_Furla	-0.59	6.42e-07
brand_MARC JACOBS (THE)	-0.73	1.69e-03
brand_MICHAEL Michael Kors	-1.72	5.78e-17
brand_Sans-Arcidet	-0.14	6.59e-01
color_Бежевый	0.12	1.54e-01
color_Голубой	0.16	2.54e-01
color_Коричневый	0.08	4.46e-01
color_Кремовый	-0.03	7.53e-01
color_Светло-бежевый	-0.00	9.95e-01
color_Черный	0.02	6.82e-01
material_Кожа натуральная: 100%	-0.02	8.12e-01
material_Текстиль: 100%	-0.37	1.25e-04
design_country_Италия	0.19	7.31e-02
design_country_США	1.28	3.82e-11
design_country_Соединенное Королевство	0.24	1.04e-01
design_country_Франция	0.31	7.88e-03
manufacture_country_Вьетнам	-0.57	2.17e-03
manufacture_country_Италия	0.92	3.87e-26
manufacture_country_Камбоджа	-0.21	2.40e-01
manufacture_country_Китай	-0.20	2.78e-02
manufacture_country_Мадагаскар	-0.90	2.03e-03
manufacture_country_Тунис	-0.22	1.64e-01
const	10.58	8.37e-198

$$R^2 = 0.690$$

У полулогарифмической модели получился очень хороший коэффициент детерминации. Модель в целом значима, и значимы оказались многие коэффициенты.

На уровне значимости 5% **отвергаем** гипотезу о значимости height, width, is\_suitcase, lock, dust\_cover, makeup\_bag, brand\_Sans-Arcidet, все цвета, material\_Кожа натуральная: 100%, design\_country\_Италия, design\_country\_Соединенное Королевство, manufacture\_country\_Камбоджа, manufacture\_country\_Тунис.

Логично, что некоторые маленькие аксессуары типа зеркала не влияют на стоимость сумки.

## 2.4 Линейная в логарифмах модель

Оценим линейную в логарифмах модель, которая объясняет логарифм цены ( $\ln(price)$ ) логарифмами размера сумки ( $\ln(height)$ ,  $\ln(width)$ ,  $\ln(depth)$ ), константой и остальными дамми-признаками.

Уравнение оцененной модели представлено в Таблице 4

Таблица 4: Уравнение полулогарифмической модели и p-value t-тестов.

признак	значение	p-value
is_suitcase	0.09	6.82e-01
keychain	0.33	4.28e-03
lock	0.50	7.15e-02
dust_cover	0.23	2.29e-01
mirror	1.00	4.87e-03
second_belt	0.31	2.96e-02
makeup_bag	0.02	8.88e-01
brand_Borbonese	-0.92	1.12e-15
brand_Bottega Veneta	0.90	1.79e-13
brand_Coccinelle	-0.62	1.92e-08
brand_Dolce & Gabbana	0.58	2.23e-06
brand_Furla	-0.58	9.85e-07
brand_MARC JACOBS (THE)	-0.71	2.10e-03
brand_MICHAEL Michael Kors	-1.72	4.44e-17
brand_Sans-Arcidet	-0.17	5.88e-01
color_Бежевый	0.11	1.85e-01
color_Голубой	0.16	2.71e-01
color_Коричневый	0.07	5.27e-01
color_Кремовый	-0.03	7.60e-01
color_Светло-бежевый	-0.01	9.54e-01
color_Черный	0.02	7.93e-01
material_Кожа натуральная: 100%	-0.00	9.92e-01
material_Текстиль: 100%	-0.36	1.83e-04
design_country_Италия	0.18	8.46e-02
design_country_США	1.29	2.04e-11
design_country_Соединенное Королевство	0.22	1.22e-01
design_country_Франция	0.31	7.22e-03
manufacture_country_Вьетнам	-0.58	1.90e-03
manufacture_country_Италия	0.94	5.23e-27
manufacture_country_Камбоджа	-0.22	2.18e-01
manufacture_country_Китай	-0.18	4.80e-02
manufacture_country_Мадагаскар	-0.87	2.65e-03
manufacture_country_Тунис	-0.23	1.36e-01
const	10.15	1.72e-118
ln_height	0.02	8.60e-01
ln_width	0.05	6.28e-01
ln_depth	0.16	1.04e-03

$$R^2 = 0.693$$

$R^2$  получился на 0.003 лучше, чем у полулогарифмической модели. Из логарифмических объясняющих переменных значимая получилась только  $\ln(\text{depth})$ . В остальном значимы все коэффициенты, кроме `is_suitcase`, `lock`, `dust_cover`, `makeup_bag`, `brand_Sans-Arcidet`, `material_Кожа натуральная`: 100%, `design_country_Италия`, `manufacture_country_Италия`, `design_country_Соединенное Королевство`, `manufacture_country_Камбоджа`, `manufacture_country_Тунис`, `цветов`, `ln_height` и `ln_width`.

## 2.5 Сравнение моделей

$R^2$  полулогарифмической модели чуть хуже, чем  $R^2$  линейной в логарифмах модели, поэтому будем сравнивать линейную модель с линейной в логарифмах тестом Бокса-Кокса с преобразованием Зарембки:

$H_0$  : «качество подгонки линейной и логарифмической моделей одинаковое»

$H_1$  : «модель с меньшей  $RSS'$  лучше»

$$\chi^2 = \frac{n}{2} \left| \ln \frac{RSS'_{\text{линейный}}}{RSS'_{\text{логарифмический}}} \right| \sim \chi_1^2$$

Где  $RSS'$  — это  $RSS$  с целевой переменной, нормированной на геометрическое среднее.

Результаты теста:

$$RSS_{\text{linear}} = 8668.18, \quad RSS_{\text{logarithmic}} = 217.77, \quad \text{p-value} = 0.0$$

Будем использовать логарифмическую модель, потому что у нее меньше  $RSS'$ , и по тесту мы отвергаем гипотезу о том, что качество моделей одинаковое.

## 3 Предпосылки теоремы Гаусса-Маркова

### 3.1 Теорема

Для начала вспомним, как звучит сама теорема Гаусса-Маркова:

В модели  $y = X\beta + \varepsilon$  оценка коэффициентов методом наименьших квадратов  $\hat{\beta} = (X'X)^{-1}X'y$  является наилучшей линейной несмещённой (BLUE), если выполнены следующие предпосылки:

#### 1. Линейность модели по параметрам

Модель линейна по параметрам  $\beta$ :  $y = X\beta + \varepsilon$

## 2. Полный ранг матрицы признаков

Матрица  $X \in \mathbb{R}^{n \times k}$  имеет полный ранг ( $\text{rank}(X) = k$ ).

Иными словами, столбцы матрицы  $X$  — линейно независимы.

## 3. Несмещённость ошибок

Ошибки имеют нулевое математическое ожидание:  $\mathbb{E}[\varepsilon] = 0$

## 4. Гомоскедастичность

Ошибки имеют одинаковую дисперсию:

$$\text{Var}(\varepsilon_i) = \sigma^2 \quad \forall i \in [1, n]$$

## 5. Отсутствие автокорреляции

Ошибки не коррелированы между собой:  $\text{Var}(\varepsilon) = \sigma^2 I_n$

Иными словами:  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0, \quad \forall i \neq j$

## 3.2 Линейная зависимость по параметрам

Проверим, что данные имеют линейную зависимость  $Y = X\beta + \varepsilon$ :

Мы задали модель, используя библиотечную функцию  $sm.OLS(y, X).fit()$ , которая строит линейную модель, подбирая параметры методом наименьших квадратов.

Таким образом, наша модель имеет вид:

$$y_i = \beta_0 + \beta_1 * x_1 + \dots + \beta_k * x_k$$

⇓

линейность по параметрам выполняется

## 3.3 Линейная независимость признаков

Для проверки линейной независимости признаков посчитаем ранг матрицы признаков и сравним его с количеством столбцов. Если значения совпадут, то матрица имеет полный ранг.

В нашем случае количество признаков (37) совпало с рангом матрицы  $\Rightarrow$  матрица имеет полный ранг, столбцы линейно независимы (таким образом, матрица обратима, МНК оценка существует).

## 3.4 Несмещенность ошибок

Проверим, что  $\mathbb{E}[\varepsilon] = 0$  (ошибки не смещены).

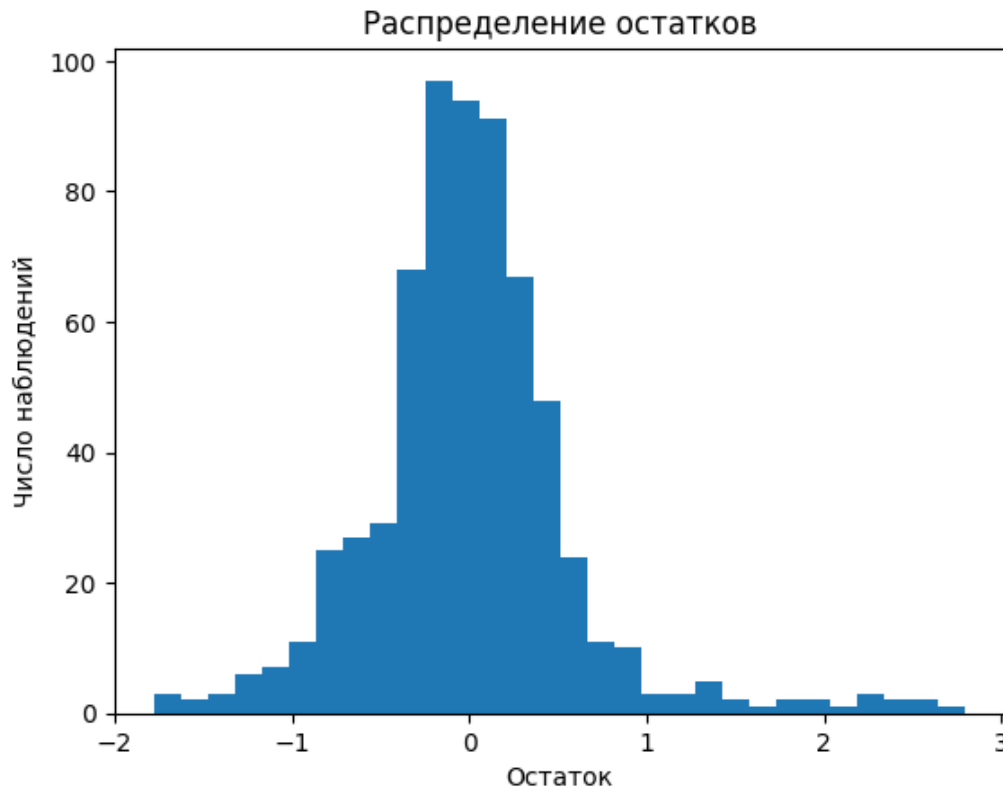
### Гипотезы:

- $H_0 : \mathbb{E}[\varepsilon] = 0$
- $H_1 : \mathbb{E}[\varepsilon] \neq 0$

Для проверки гипотезы используем обычный t-тест на 95% уровне значимости.

После проведения теста получили, что значение t-статистики равно 0, а p-значение – 1. Таким образом, у нас нет оснований для отвержения основной гипотезы  $\Rightarrow \mathbb{E}[\varepsilon] = 0$

Для наглядности визуализируем распределение ошибок:



## 3.5 Гомоскедастичность

Для решения этой задачи используем тест Бреуша–Пагана:

### Гипотезы:

- $H_0 : \text{Var}(\varepsilon_i) = \sigma^2$  (гомоскедастичность)
- $H_1 : \text{Var}(\varepsilon_i) = \sigma^2 h(z_i)$ ,  $h(z_i)$ ,  $z_i$  — переменные, влияющие на дисперсию ошибок. Иными словами, присутствует гетероскедастичность

### Идея теста:

Квадраты остатков  $\hat{\varepsilon}_i^2$  и регрессируются на признаки модели:  $\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 z_{i1} + \dots + \alpha_k z_{ik} + u_i$ .

### Статистика теста:

$LM = n \cdot R^2$  – сравнивается с критическим значением распределения  $\chi_k^2$ .

После проведения теста получили, значение  $LM = 143.8$ , при критическом значении  $\chi_{\alpha=0.05}^2 = 23.5$ . Так как  $143.8 > 23.5 \Rightarrow H_0$  отвергается.

График ниже также иллюстрирует наличие гетероскедастичности:



Поскольку тест Бреуша–Пагана показал наличие гетероскедастичности, использование обычных стандартных ошибок может привести к некорректным статистическим выводам. Для устранения этой проблемы в дальнейших расчётах применяются робастные оценки стандартных ошибок, устойчивые к неоднородности дисперсии. Эти оценки позволяют корректно интерпретировать  $t$ -статистики и  $p$ -value, даже при нарушении предпосылки о гомоскедастичности ошибок.

## 3.6 Некоррелируемость ошибок

Я устала. Для проверки этого факта используем тест Дарбина-Уотсона.

### Гипотезы:

- $H_0$  : ошибки не автокоррелированы (независимы)
- $H_1$ : ошибки автокоррелированы (зависят от предыдущих)

**Статистика теста:**

$$DW = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2}$$

**Идея теста:**

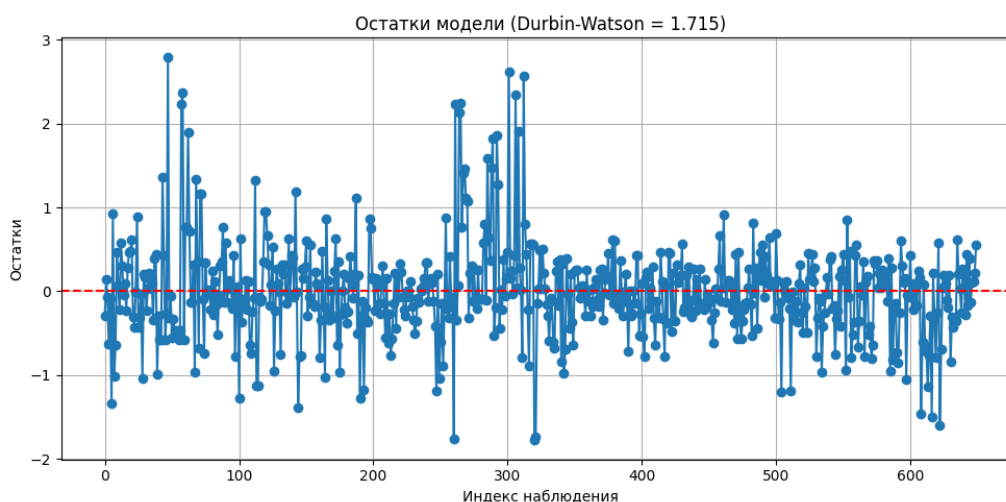
Статистика показывает, насколько ошибки "прыгают" между наблюдениями

Значения теста соответствуют следующим интерпретациям:

Значение $DW$	Интерпретация
$\approx 2$	Нет автокорреляции
$< 2$	Положительная автокорреляция
$> 2$	Отрицательная автокорреляция

Таблица 5: Интерпретация статистики Дарбина-Уотсона

Визуализация остатков модели для каждого из наблюдений:



Значение DW-статистики составило 1.7, что соответствует умеренной положительной автокорреляции. Аналогично прошлому пункту, для корректной оценки стандартных ошибок применим робастные оценки, устойчивые к внутригрупповой корреляции.

## 4 Гипотеза о нормальности случайной ошибки

Проведем тест Колмогорова-Смирнова на нормальность распределения случайной ошибки модели



## Основная гипотеза:

- $H_0$ : Случайная ошибка происходит из нормального распределения с заданными параметрами
- $H_1$ : Случайная ошибка *не* происходит из нормального распределения

## Статистика теста:

Статистика Колмогорова-Смирнова (K-S) вычисляется как максимальное расхождение между эмпирической функцией распределения (ЭФР) и теоретической нормальной функцией распределения:

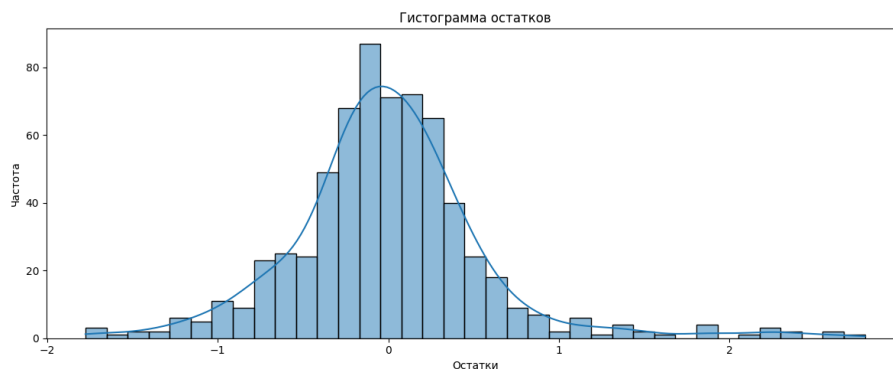
$$D_n = \sup_x |F_n(x) - F(x)|$$

Вот результаты теста:

Тест Колмогорова Смирнова на нормальность распределения:  
Kolmogorov-Smirnov test: D = 0.0933, p-value = 0.0000  
Гипотеза о нормальности отвергается

Получается гипотеза о нормальности отвергается на любом уровне значимости, так как p-value < 0.01.

Нарисуем еще гистограмму остатков:



## 5 Гипотеза о значимости коэффициентов и модели в целом

### 5.1 Значимость коэффициентов

Для оценки значимости параметров был повсеместно использован t-тест:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

$$t_i = \frac{\hat{\beta}_i}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_i)}} \sim t_{n-k}$$

## 5.2 Значимость модели

Для проверки значимости модели в целом был использован F-test:

$$H_0 : \beta_i = 0 \quad \forall i \geq 1$$

$$H_1 : \text{«хотя бы одно из ограничений неверно.»}$$

При верной  $H_0$ :

$$F_{H_0} = \frac{R_{UR}^2 / (k_{UR} - 1)}{(1 - R_{UR}^2) / (n - k_{UR})} \sim F(k_{UR} - 1, n - k_{UR})$$

$$\mathcal{K}_\alpha = (F_{k-1, n-k, 1-\alpha}, \infty)$$

$$\text{p-value} = 1 - F_{F_{k-1, n-k}}(F_{obs})$$

## 6 Интерпретация итоговой модели

### Аксессуары и дополнительные элементы

- **Наибольшее повышение цены:**

- Зеркало:  $e^{1.004750} \approx 2.73$  раза (+173%)
- Замок:  $e^{0.5031} \approx 1.65$  раза (+65%)
- Второй ремень:  $e^{0.309583} \approx 1.36$  раза (+36%)
- Брелок:  $e^{0.326285} \approx 1.38$  раза (+38%)

- **Слабое влияние:**

- Косметичка:  $e^{0.0168} \approx 1.02$  раза (+2%)
- Пыльник:  $e^{0.2315} \approx 1.26$  раза (+26%)
- Чемодан:  $e^{0.0901} \approx 1.09$  раза (+9%)

### Влияние брендов

- **Повышают цену:**

- Bottega Veneta:  $e^{0.901444} \approx 2.46$  раза (+146%)
- Dolce & Gabbana:  $e^{0.578613} \approx 1.78$  раза (+78%)

- **Понижают:**

- MICHAEL Michael Kors:  $e^{-1.721312} \approx 0.18$  раза (-82%)
- Borbonese:  $e^{-0.921914} \approx 0.40$  раза (-60%)
- MARC JACOBS:  $e^{-0.714986} \approx 0.49$  раза (-51%)

## Влияние цвета

- **Повышают цену:**

- Голубой:  $e^{0.156121} \approx 1.17$  раза (+17%)
- Бежевый:  $e^{0.109365} \approx 1.12$  раза (+12%)
- Коричневый:  $e^{0.067140} \approx 1.07$  раза (+7%)

- **Незначительное влияние:**

- Черный:  $e^{0.015548} \approx 1.02$  раза (+2%)
- Кремовый:  $e^{-0.032103} \approx 0.97$  раза (-3%)
- Светло-бежевый:  $e^{-0.007977} \approx 0.99$  раза (-1%)

## Влияние материалов

- Текстиль:  $e^{-0.355305} \approx 0.70$  раза (-30%)
- Натуральная кожа:  $e^{-0.000604} \approx 1.00$  раза (без влияния)

## Влияние страны производства и дизайна

- **Дизайн:**

- США:  $e^{1.290369} \approx 3.63$  раза (+263%)
- Франция:  $e^{0.308532} \approx 1.36$  раза (+36%)
- Италия:  $e^{0.182081} \approx 1.20$  раза (+20%)

- **Производство:**

- Италия:  $e^{0.936172} \approx 2.55$  раза (+155%)
- Китай:  $e^{-0.183101} \approx 0.83$  раза (-17%)
- Вьетнам:  $e^{-0.578694} \approx 0.56$  раза (-44%)

## Базовые параметры

- Константа (базовая цена):  $e^{10.148609} \approx 25\,570$ .

## Вывод

На цену наиболее значительно влияют:

- Премиальные бренды (Bottega Veneta, Dolce & Gabbana)
- Страна дизайна (США, Франция) и производства (Италия)
- Наличие премиальных аксессуаров (зеркало, замок)

## 7 Прогноз цены своего товара

### 7.1 Предсказание на реальных данных

Рассмотрим [THE TOTE BAG](#), у которой не были указаны размеры, из-за чего она не попала в итоговый датасет. Попробуем предсказать для нее цену, используя средние лог размеры.

Получаем 46 тыс., что чуть выше оригинальной цены без скидки в 41 тыс., ура!

### 7.2 Предсказание на смоделированных данных

Также попробуем использовать модель для предсказания цены авторской сумки<sup>1</sup>, которой еще нет в продаже. Очень вероятно, что эта сумка будет выставлена на аукционе и, чтобы суметь заработать на этом событии, необходимо знать ее справедливую стоимость. Вот тут наша модель и пригодится!



Рис. 1: Удивительной красоты сумка

Цена этой сумки получилась  $e^{13.03} = 455887$  рублей :) Теперь ритейлеры сумок могут понимать адекватную цену за этот товар и в случае покупки за цену ниже данной, продать ее коллекционерам.

---

<sup>1</sup>Дизайнером сумки является один из авторов проекта

# 8 Квантильная регрессия

Результаты квантильной регрессии для разных квантилей (0.1, 0.5, 0.75, 0.9)

=== Результаты квантильной регрессии для квантиля 0.1 ===										
QuantReg Regression Results										
Dep. Variable:	price	Pseudo R-squared:	0.4664							
Model:	QuantReg	Bandwidth:	0.2091							
Method:	Least Squares	Sparsity:	1.244							
Date:	Thu, 08 May 2025	No. Observations:	650							
Time:	20:53:43	Df Residuals:	613							
		Df Model:	36							
=====										
	coef	std err	t	P> t	[0.025	0.975]				
-----										
is_suitcase	-0.3267	0.138	-2.363	0.018	-0.598	-0.055				
keychain	0.0208	0.071	0.403	0.687	-0.112	0.169				
lock	0.1333	0.175	0.750	0.454	-0.213	0.475				
dust_cover	0.0673	0.121	0.558	0.577	-0.170	0.304				
mirror	1.2667	0.223	5.671	0.000	0.828	1.705				
second_belt	0.2676	0.089	3.000	0.003	0.092	0.443				
makeup_bag	0.0328	0.075	0.437	0.662	-0.114	0.180				
brand_Borbonese	-0.4772	0.070	-6.781	0.000	-0.615	-0.339				
brand_Bottega Veneta	1.1001	0.075	14.712	0.000	0.958	1.254				
brand_Coccinelle	-0.6001	0.069	-9.915	0.000	-0.815	-0.545				
brand_Dolce & Gabbana	0.8871	0.076	11.657	0.000	0.738	1.037				
brand_Furla	-0.3970	0.073	-5.402	0.000	-0.541	-0.253				
brand_MARC JACOBS (THE)	-0.1366	0.145	-0.939	0.348	-0.422	0.149				
brand_MICHAEL Michael Kors	-0.7057	0.125	-6.126	0.000	-1.011	-0.520				
brand_Sans-Arcidet	-0.4325	0.192	-2.251	0.025	-0.810	-0.055				
color_бежевый	0.1232	0.052	2.382	0.018	0.022	0.225				
color_голубой	0.1576	0.089	1.771	0.077	-0.017	0.332				
color_коричневый	0.2094	0.067	3.140	0.002	0.078	0.340				
color_красный	0.1207	0.066	1.831	0.068	-0.009	0.250				
color_Светло-бежевый	-0.0301	0.088	-0.435	0.664	-0.218	0.134				
color_Черный	0.0242	0.037	0.651	0.516	-0.049	0.097				
material_Кожа натуральная: 100%	0.2959	0.040	7.474	0.000	0.218	0.374				
material_Текстиль: 100%	-0.2000	0.059	-3.374	0.001	-0.316	-0.084				
design_country_Италия	0.3201	0.066	4.834	0.000	0.190	0.450				
design_country_США	0.5901	0.119	4.972	0.000	0.357	0.823				
design_country_Соединенное Королевство	0.5473	0.090	6.049	0.000	0.370	0.725				
design_country_Франция	0.4342	0.072	6.038	0.000	0.293	0.575				
manufacture_country_Вьетнам	-0.0031	0.117	-0.713	0.475	-0.312	0.146				
manufacture_country_Италия	0.5391	0.052	10.384	0.000	0.437	0.641				
manufacture_country_Камбожа	0.0784	0.114	0.687	0.492	-0.146	0.303				
manufacture_country_Китай (Китайская Народная Республика)	-0.0604	0.058	-1.039	0.299	-0.174	0.054				
manufacture_country_Малагаскар	-0.3781	0.101	-2.094	0.037	-0.573	-0.023				
manufacture_country_Тунис	-0.1088	0.098	-1.113	0.266	-0.301	0.083				
const	8.7577	0.218	40.202	0.000	8.330	9.186				
ln_height	0.1601	0.059	2.820	0.005	0.050	0.282				
ln_width	0.2278	0.069	3.294	0.001	0.092	0.364				
ln_depth	0.0549	0.030	1.802	0.072	-0.005	0.115				
=====										
=== Результаты квантильной регрессии для квантиля 0.5 ===										
QuantReg Regression Results										
Dep. Variable:	price	Pseudo R-squared:	0.5773							
Model:	QuantReg	Bandwidth:	0.1993							
Method:	Least Squares	Sparsity:	0.9253							
Date:	Thu, 08 May 2025	No. Observations:	650							
Time:	21:01:05	Df Residuals:	613							
		Df Model:	36							
=====										
	coef	std err	t	P> t	[0.025	0.975]				
-----										
is_suitcase	0.7879	0.148	5.309	0.000	0.496	1.079				
keychain	0.2275	0.077	2.964	0.003	0.077	0.378				
lock	0.2329	0.100	1.239	0.216	-0.195	0.602				
dust_cover	0.4647	0.130	3.584	0.000	0.210	0.719				
mirror	1.6916	0.240	7.054	0.000	1.221	2.163				
second_belt	0.1601	0.096	1.672	0.095	-0.028	0.348				
makeup_bag	-0.0100	0.081	-0.236	0.813	-0.177	0.159				
brand_Borbonese	-1.1261	0.076	-14.906	0.000	-1.274	-0.978				
brand_Bottega Veneta	0.6471	0.081	8.016	0.000	0.489	0.806				
brand_Coccinelle	-0.6504	0.074	-8.832	0.000	-0.795	-0.506				
brand_Dolce & Gabbana	0.3157	0.082	3.863	0.000	0.155	0.476				
brand_Furla	-0.4273	0.079	-5.416	0.000	-0.582	-0.272				
brand_MARC JACOBS (THE)	-1.4629	0.156	-9.372	0.000	-1.769	-1.156				
brand_MICHAEL Michael Kors	-2.2452	0.134	-16.732	0.000	-2.509	-1.982				
brand_Sans-Arcidet	-0.5877	0.206	-2.849	0.005	-0.993	-0.183				
color_бежевый	0.0543	0.056	0.977	0.329	-0.055	0.163				
color_голубой	0.0113	0.096	0.118	0.906	-0.176	0.199				
color_коричневый	-0.0103	0.072	-0.144	0.885	-0.151	0.130				
color_красный	-0.1152	0.071	-1.627	0.104	-0.254	0.024				
color_Светло-бежевый	0.1415	0.094	1.505	0.133	-0.043	0.326				
color_Черный	8.267e-09	0.040	2.07e-07	1.000	-0.079	0.079				
material_Кожа натуральная: 100%	0.0809	0.043	1.904	0.057	-0.003	0.164				
material_Текстиль: 100%	-0.1768	0.064	-2.777	0.006	-0.302	-0.052				
design_country_Италия	-0.0600	0.071	-0.844	0.399	-0.200	0.080				
design_country_США	1.5926	0.127	12.499	0.000	1.342	1.843				
design_country_Соединенное Королевство	-0.1500	0.057	-1.637	0.102	-0.350	0.012				
design_country_Франция	0.1085	0.077	1.406	0.160	-0.043	0.260				
manufacture_country_Вьетнам	-0.3606	0.125	-2.882	0.004	-0.606	-0.115				
manufacture_country_Италия	1.1145	0.056	19.936	0.000	1.005	1.224				
manufacture_country_Камбожа	-0.1158	0.102	-0.977	0.323	-0.361	0.121				
manufacture_country_Китай (Китайская Народная Республика)	-0.1302	0.062	-2.089	0.037	-0.253	-0.008				
manufacture_country_Малагаскар	-0.5036	0.104	-2.597	0.010	-0.884	-0.123				
manufacture_country_Тунис	-0.2620	0.105	-2.496	0.013	-0.468	-0.056				
const	10.2736	0.234	43.926	0.000	9.814	10.733				
ln_height	0.1495	0.063	2.363	0.018	0.025	0.274				
ln_width	-0.0243	0.074	-0.327	0.744	-0.170	0.122				
ln_depth	0.1114	0.033	3.468	0.001	0.040	0.178				
=====										
=== Результаты квантильной регрессии для квантиля 0.9 ===										
QuantReg Regression Results										
Dep. Variable:	price	Pseudo R-squared:	0.5175							
Model:	QuantReg	Bandwidth:	0.1938							
Method:	Least Squares	Sparsity:	0.7149							
Date:	Thu, 08 May 2025	No. Observations:	650							
Time:	20:56:09	Df Residuals:	613							
		Df Model:	36							
=====										
	coef	std err	t	P> t	[0.025	0.975]				
-----										
is_suitcase	-0.1683	0.132	-1.271	0.204	-0.428	0.092				
keychain	0.1297	0.068	1.894	0.059	-0.005	0.264				
lock	0.3405	0.168	2.031	0.043	0.011	0.670				
dust_cover	0.2633	0.116	2.276	0.023	0.036	0.490				
mirror	0.8064	0.214	3.769	0.000	0.386	1.227				
second_belt	0.1813	0.085	2.122	0.034	0.014	0.349				
makeup_bag	-0.0537	0.072	-0.747	0.455	-0.195	0.087				
brand_Borbonese	-0.6587	0.067	-7.773	0.000	-0.791	-0.526				
brand_Bottega Veneta	1.1282	0.072	15.666	0.000	0.987	1.270				
brand_Coccinelle	-0.4156	0.066	-6.325	0.000	-0.545	-0.287				
brand_Dolce & Gabbana	0.6705	0.073	9.187	0.000	0.527	0.814				
brand_Furla	-0.2304	0.070	-3.273	0.001	-0.369	-0.092				
brand_MARC JACOBS (THE)	-0.2983	0.139	-2.142	0.033	-0.572	-0.025				
brand_MICHAEL Michael Kors	-1.7054	0.120	-11.400	0.000	-1.601	-1.120				
brand_Sans-Arcidet	-0.7552	0.184	-4.104	0.000	-1.117	-0.394				
color_бежевый	0.1012	0.050	2.042	0.042	0.004	0.199				
color_голубой	-0.0472	0.085	-0.553	0.580	-0.215	0.120				
color_коричневый	0.0684	0.064	1.073	0.285	-0.057	0.184				
color_красный	-0.0017	0.063	-0.028	0.978	-0.126	0.122				
color_Светло-бежевый	0.0203	0.084	0.242	0.809	-0.144	0.185				
color_Черный	1.569e-06	0.035	4.394e-05	1.000	-0.070	0.070				
material_Кожа натуральная: 100%	0.0927	0.038	2.446	0.015	0.018	0.166				
material_Текстиль: 100%	-0.1190	0.057	-2.095	0.037	-0.231	-0.007				
design_country_Италия	-0.0896	0.063	-1.413	0.158	-0.214	0.063				
design_country_США	0.8386	0.114	7.377	0.000	0.615	1.062				
design_country_Соединенное Королевство	0.8119	0.097	8.187	0.000	0.659	0.959				
design_country_Франция	0.3410	0.069	4.951	0.000	0.206	0.476				
manufacture_country_Вьетнам	-0.5579	0.112	-4.997	0.000	-0.777	-0.339				
manufacture_country_Италия	0.8865	0.050	17.725	0.000	0.787	0.987				
manufacture_country_Камбожа	-0.0324	0.109	-0.296	0.767	-0.247	0.182				
manufacture_country_Китай (Китайская Народная Республика)	-0.1178	0.056	-2.117	0.033	-0.227	-0.089				
manufacture_country_Малагаскар	-0.5282	0.173	-3.053	0.002	-0.868	-0.188				
manufacture_country_Тунис	-0.1316	0.094	-1.427	0.154	-0.319	0.067				
const	9.5462	0.209	45.479	0.000	8.900	10.196				
ln_height	0.1622	0.056	2.875	0.004	0.051	0.273				
ln_width	0.0915	0.060	1.508	0.130	-0.122	0.313				
ln_depth	0.1710	0.029	5.864	0.000	0.114	0.228				
=====										

- Для  $q=0.5$  (медиана):

- Pseudo  $R^2 = 0.518$  (лучше объясняющая способность)
- Ключевые драйверы: производство в Италии, бренд Bottega Veneta, натуральная кожа

- Для  $q=0.9$  (верхний квантиль):

- Pseudo  $R^2 = 0.570$  (наивысший)
- Сильнее всего влияют: производство в Италии, текстильный материал, бренд Borbonese

## Наблюдаемые закономерности

### Бренды

- Bottega Veneta стабильно повышает цену на всех квантилях
- Эффект бренда ослабевает для премиум-сегмента ( $q=0.9$ ) у Bottega Veneta, Dolce & Gabbana

### Материалы

- Натуральная кожа:
  - Положительный эффект снижается с ростом квантиля
  - Для дорогих сумок ( $q=0.9$ ) становится негативным фактором

### Страна производства

- Италия:
  - Максимальный эффект для премиума (для  $q=0.9$ )
- Китай/Вьетнам:
  - Негативное влияние усиливается для медианы ( $q=0.5$ )

## Неожиданные эффекты

- Чемоданные модели (is\_suitcase):

- Снижают цену на нижних квантилях
- Повышают цену для премиум-сегмента ( $q=0.9$ )

- Размеры:

- Высота (ln\_height) значима для  $q=0.1-0.5$ , но не для  $q=0.9$
- Глубина (ln\_depth) значима на всех квантилях

=====

=== Результаты квантильной регрессии для квантиля 0.1 (бутстрап, n=100) ===

=====

	factor	coeff	std err (boot)	95% CI low	95% CI high	t-stat	p-value
0	is_suitcase	-0.3267	0.2556	-0.8276	0.1743	-1.2781	0.2012
1	keychain	0.0288	0.2326	-0.4270	0.4847	0.1239	0.9014
2	lock	0.1313	0.2953	-0.4475	0.7100	0.4446	0.6566
3	dust_cover	0.0673	0.1583	-0.2428	0.3775	0.4255	0.6705
4	mirror	1.2667	0.5607	0.1678	2.3656	2.2593	0.0239
5	second_belt	0.2676	0.2191	-0.1619	0.6971	1.2211	0.2320
6	makeup_bag	0.0328	0.1313	-0.2246	0.2901	0.2496	0.8029
7	brand_Borbonese	-0.4772	0.1191	-0.7107	-0.2436	-0.4048	0.0001
8	brand_Bottega Veneta	1.1061	0.1227	0.8656	1.3465	9.0154	0.0000
9	brand_Coccinelle	-0.6801	0.0819	-0.8407	-0.5195	-8.3016	0.0000
10	brand_Dolce & Gabbana	0.8871	0.1545	0.5843	1.1899	5.7425	0.0000
11	brand_Furla	-0.3970	0.1067	-0.6060	-0.1879	-3.7218	0.0002
12	brand_MARC JACOBS (THE)	-0.1366	0.3378	-0.7987	0.5256	-0.4042	0.6861
13	brand_MICHAEL Michael Kors	-0.7657	0.2593	-1.2738	-0.2575	-2.9533	0.0031
14	brand_Sans-Arcidet	-0.4225	0.2708	-0.9633	0.0983	-1.5970	0.1103
15	color_Белый	0.1232	0.0670	-0.0081	0.2546	1.8308	0.0659
16	color_Голубой	0.1576	0.1028	-0.0439	0.3591	1.5332	0.1252
17	color_Коричневый	0.2094	0.0874	0.0381	0.3807	2.3964	0.0166
18	color_Кремлевый	0.1207	0.1982	-0.2677	0.5091	0.6092	0.5424
19	color_Светло-бежевый	-0.0381	0.1283	-0.2896	0.2135	-0.2966	0.7667
20	color_Черный	0.0242	0.0392	-0.0525	0.1010	0.6189	0.5360
21	material_Кожа натуральная: 100%	0.2959	0.0777	0.1435	0.4482	3.8005	0.0001
22	material_Текстиль: 100%	-0.2000	0.1709	-0.5349	0.1349	-1.1706	0.2417
23	design_country_Италия	0.1021	0.2899	-0.2481	0.8883	1.1042	0.2695
24	design_country_США	0.5901	0.4344	-0.2615	1.4416	1.3582	0.1744
25	design_country_Соединенное Королевство	0.5473	0.2885	-0.0181	1.1127	1.8973	0.0578
26	design_country_Франция	0.4342	0.3339	-0.2202	1.0887	1.3004	0.1934
27	manufacture_country_Вьетнам	-0.0631	0.3335	-0.7367	0.5705	-0.2491	0.8003
28	manufacture_country_Италия	0.5191	0.1265	0.2911	0.7471	4.2601	0.0000
29	manufacture_country_Каюбожа	0.0784	0.1967	-0.3071	0.4640	0.3987	0.6901
30	manufacture_country_Китай (Китайская Народная ...	-0.0604	0.1317	-0.3185	0.1977	-0.4584	0.6467
31	manufacture_country_Мадагаскар	-0.3781	0.2387	-0.8461	0.0898	-1.5838	0.1132
32	manufacture_country_Тунис	-0.0808	0.1790	-0.4597	0.2421	-0.6077	0.5434
33	const	8.7577	0.4177	7.9390	9.5764	20.9659	0.0000
34	ln_height	0.1661	0.1437	-0.1155	0.4478	1.1562	0.2476
35	ln_width	0.2278	0.1351	-0.0370	0.4926	1.6862	0.0918
36	ln_depth	0.0549	0.0672	-0.0769	0.1867	0.8158	0.4146

=====

=== Результаты квантильной регрессии для квантиля 0.75 (бутстрап, n=100) ===

=====

	factor	coeff	std err (boot)	95% CI low	95% CI high	t-stat	p-value
0	is_suitcase	0.7879	0.5902	-0.3689	1.9446	1.3350	0.1819
1	keychain	0.2275	0.1691	-0.1039	0.5589	1.3457	0.1784
2	lock	0.2329	0.9877	-1.7029	2.1687	0.2358	0.8136
3	dust_cover	0.4647	0.2277	0.0185	0.9109	2.0411	0.0412
4	mirror	1.6916	0.6528	0.4122	2.9711	2.5914	0.0096
5	second_belt	0.1601	0.0864	-0.0093	0.3295	1.8523	0.0640
6	makeup_bag	-0.0190	0.0860	-0.1875	0.1495	-0.2211	0.8250
7	brand_Borbonese	-1.1261	0.1405	-1.4015	-0.8506	-8.0126	0.0000
8	brand_Bottega Veneta	0.6471	0.2147	0.2262	1.0679	3.0134	0.0026
9	brand_Coccinelle	-0.6504	0.1084	-0.8628	-0.4381	-6.0030	0.0000
10	brand_Dolce & Gabbana	0.3157	0.1640	-0.0059	0.6372	1.9242	0.0543
11	brand_Furla	-0.4273	0.1261	-0.6744	-0.1803	-3.3899	0.0007
12	brand_MARC JACOBS (THE)	-1.4629	0.5409	-2.5232	-0.4027	-2.7044	0.0068
13	brand_MICHAEL Michael Kors	-2.2452	0.4688	-3.1641	-1.3264	-4.7894	0.0002
14	brand_Sans-Arcidet	-0.5877	0.3425	-1.2590	0.0837	-1.7157	0.0862
15	color_Белый	0.0543	0.0622	-0.0676	0.1761	0.8725	0.3829
16	color_Голубой	0.0113	0.2833	-0.5439	0.5665	0.0399	0.9681
17	color_Коричневый	-0.0103	0.0920	-0.1906	0.1700	-0.1123	0.9106
18	color_Кремлевый	-0.1152	0.0812	-0.2744	0.0440	-1.4179	0.1562
19	color_Светло-бежевый	0.1415	0.1376	-0.1281	0.4112	1.0287	0.3036
20	color_Черный	0.0000	0.0536	-0.1051	0.1051	0.0000	1.0000
21	material_Кожа натуральная: 100%	0.0609	0.0806	-0.0771	0.2389	1.0040	0.3154
22	material_Текстиль: 100%	-0.1768	0.1312	-0.4340	0.0804	-1.3471	0.1780
23	design_country_Италия	-0.0600	0.1500	-0.3541	0.2341	-0.3999	0.6893
24	design_country_США	1.5926	0.4785	0.6547	2.5305	3.3282	0.0009
25	design_country_Соединенное Королевство	-0.1590	0.1638	-0.4801	0.1621	-0.9704	0.3318
26	design_country_Франция	0.1085	0.1469	-0.1795	0.3965	0.7385	0.4602
27	manufacture_country_Вьетнам	-0.3606	0.3498	-1.0462	0.3250	-1.0308	0.3026
28	manufacture_country_Италия	1.1145	0.1110	0.8969	1.3320	10.0418	0.0000
29	manufacture_country_Каюбожа	-0.1198	0.1363	-0.3870	0.1473	-0.8790	0.3794
30	manufacture_country_Китай (Китайская Народная ...	-0.1302	0.0799	-0.2869	0.0264	-1.6292	0.1033
31	manufacture_country_Мадагаскар	-0.5036	0.2968	-1.0852	0.0781	-1.6968	0.0897
32	manufacture_country_Тунис	-0.2620	0.1348	-0.5263	0.0022	-1.9433	0.0520
33	const	10.2736	0.4069	9.4762	11.0711	25.2505	0.0000
34	ln_height	0.1495	0.0816	-0.0104	0.3093	1.8326	0.0669
35	ln_width	-0.0243	0.1027	-0.2256	0.1770	-0.2363	0.8132
36	ln_depth	0.1134	0.0480	0.0193	0.2074	2.3614	0.0182

=====

=== Результаты квантильной регрессии для квантиля 0.9 (бутстрап, n=100) ===

=====

	factor	coeff	std err (boot)	95% CI low	95% CI high	t-stat	p-value
0	is_suitcase	-0.1683	0.6800	-1.5010	1.1645	-0.2474	0.8046
1	keychain	0.1297	0.1145	-0.0948	0.3542	1.1325	0.2574
2	lock	0.3405	0.5651	-0.7071	1.4482	0.6026	0.5468
3	dust_cover	0.2693	0.1656	-0.0613	0.5979	1.5996	0.1119
4	mirror	0.0664	0.3027	-0.1789	1.7917	0.2041	0.1087
5	second_belt	0.1813	0.0755	0.0334	0.3293	2.4018	0.0163
6	makeup_bag	-0.0537	0.0732	-0.1972	0.0898	-0.7331	0.4635
7	brand_Borbonese	-0.6587	0.1152	-0.8844	-0.4329	-5.7193	0.0000
8	brand_Bottega Veneta	1.1282	0.1315	0.8704	1.3859	8.5791	0.0000
9	brand_Coccinelle	-0.4156	0.0670	-0.5468	-0.2843	-6.2067	0.0000
10	brand_Dolce & Gabbana	0.6705	0.1224	0.4305	0.9105	5.4760	0.0000
11	brand_Furla	-0.2304	0.0863	-0.3996	-0.0611	-2.6682	0.0076
12	brand_MARC JACOBS (THE)	-0.2983	0.7587	-1.7854	1.1888	-0.3931	0.6942
13	brand_MICHAEL Michael Kors	-1.3654	0.7988	-2.9311	0.2002	-1.7093	0.0874
14	brand_Sans-Arcidet	-0.7552	0.2252	-1.1966	-0.3137	-3.3526	0.0008
15	color_Белый	0.3012	0.0824	-0.0212	0.2236	1.6203	0.1052
16	color_Голубой	-0.0472	0.0884	-0.2204	0.1261	-0.5336	0.5936
17	color_Коричневый	0.0584	0.0715	-0.0717	0.2085	0.9573	0.3384
18	color_Кремлевый	-0.0017	0.0553	-0.1101	0.1067	-0.0315	0.9749
19	color_Светло-бежевый	0.0203	0.1262	-0.2270	0.2677	0.1612	0.8719
20	color_Черный	0.0000	0.0299	-0.0586	0.0586	0.0001	1.0000
21	material_Кожа натуральная: 100%	0.0927	0.0620	-0.0288	0.2143	1.4956	0.1348
22	material_Текстиль: 100%	-0.1190	0.1101	-0.3348	0.0968	-1.0811	0.2797
23	design_country_Италия	-0.0896	0.1073	-0.2999	0.1206	-0.8357	0.4033
24	design_country_США	0.8386	0.7949	-0.7195	2.3967	1.0459	0.2915
25	design_country_Соединенное Королевство	0.1115	0.1178	-0.1193	0.3424	0.9469	0.3482
26	design_country_Франция	0.3410	0.1519	0.0434	0.6386	2.2456	0.0247
27	manufacture_country_Вьетнам	-0.5379	0.4428	-1.4237	0.3100	-1.2599	0.2077
28	manufacture_country_Италия	0.8886	0.1161	0.6612	1.1161	7.6570	0.0000
29	manufacture_country_Каюбожа	-0.0324	0.1384	-0.3036	0.2389	-0.2338	0.8152
30	manufacture_country_Китай (Китайская Народная ...	-0.1178	0.0752	-0.2651	0.0296	-1.5667	0.1172
31	manufacture_country_Мадагаскар	-0.5282	0.2046	-0.9293	-0.1271	-2.5810	0.0099
32	manufacture_country_Тунис	-0.1336	0.1134	-0.3560	0.0887	-1.1671	0.2388
33	const	9.5462	0.3445	8.8710	10.2215	27.7072	0.0000
34	ln_height	0.1622	0.0711	0.0228	0.3016	2.2813	0.0225
35	ln_width	0.0915	0.0893	-0.0835	0.2666	1.0247	0.3055
36	ln_depth	0.1710	0.0483	0.0763	0.2657	3.5395	0.0004

=====

=== Результаты квантильной регрессии для квантиля 0.9 (бутстрап, n=100) ===

=====

	factor	coeff	std err (boot)	95% CI low	95% CI high	t-stat	p-value
0	is_suitcase	0.5658	0.4552	-0.3265	1.4581	1.2429	0.2139
1	keychain	0.6794	0.3485	-0.0036	1.3624	1.9496	0.0512
2	lock	0.5981	0.7611	-0.8937	2.0899	0.7859	0.4320
3	dust_cover	0.4381	0.3256	-0.2000	1.0762	1.3458	0.1784
4	mirror	0.8036	0.5231	-0.2217	1.8290	1.5362	0.1245
5	second_belt	0.0959	0.2989	-0.4900	0.6818	0.3208	0.7484
6	makeup_bag	-0.0701	0.1051	-0.2761	0.1360	-0.6664	0.5052
7	brand_Borbonese	-1.3849	0.2373	-1.8501	-0.9198	-5.8356	0.0000
8	brand_Bottega Veneta	0.5522	0.4426	-0.3153	1.4197	1.2477	0.2122
9	brand_Coccinelle	-0.7085	0.1337	-0.9704	-0.4465	-5.3010	0.0000
10	brand_Dolce & Gabbana	0.1622	0.4475	-0.7149	1.0393	0.3625	0.7170
11	brand_Furla	-1.0778	0.4063	-1.8742	-0.2815	-2.6528	0.0080
12	brand_MARC JACOBS (THE)	-0.5181	0.6157	-1.7249	0.6887	-0.8415	0.4001
13	brand_MICHAEL Michael Kors	-1.3332	0.5625	-2.4357	-0.2307	-2.3702	0.0178
14	brand_Sans-Arcidet	0.4475	0.4783	-0.4900	1.3849	0.9356	0.3495
15	color_Белый	0.0075	0.0637	-0.1174	0.1325	0.1180	0.9060
16	color_Голубой	0.0119	0.4253	-0.8217	0.8455	0.0280	0.9777
17	color_Коричневый	0.0075	0.1114	-0.2109	0.2259	0.0675	0.9462
18	color_Кремлевый	-0.1759	0.1515	-0.4728	0.1210	-1.1612	0.2455
19	color_Светло-бежевый	0.0601	0.1795	-0.2917	0.4120	0.3350	0.7376
20	color_Черный	0.0075	0.0671	-0.1241	0.1391	0.1120	0.9108
21	material_Кожа натуральная: 100%	-0.6146	0.4211	-1.4399	0.2107	-1.4595	0.1444
22	material_Текстиль: 100%	-0.8423	0.4270	-1.6792	-0.0054	-1.9727	0.0485
23	design_country_Италия	-0.0297	0.1239	-0.2724	0.2131	-0.2396	0.8106
24	design_country_США	-0.0227	0.5573	-0.4696	1.7150	0.1133	0.2639
25	design_country_Соединенное Королевство	-0.0263	0.1765	-0.3722	0.3195	-0.1492	0.8814
26	design_country_Франция	0.0561	0.1494	-0.2367	0.3489	0.3755	0.7073
27	manufacture_country_Вьетнам	-0.3940	0.2307	-0.8462	0.0582	-1.7077	0.0877
28	manufacture_country_Италия	1.3883	0.1223	1.1487	1.6279	11.3548	0.0000
29	manufacture_country_Камбоджа	-0.4232	0.2680	-0.9484	0.1020	-1.5793	0.1143
30	manufacture_country_Китай (Китайская Народная Республика)	-0.0644	0.1218	-0.3032	0.1744	-0.5288	0.5970
31	manufacture_country_Малайзия	-1.8216	0.1616	-3.0305	-0.6126	-2.9532	0.0031
32	manufacture_country_Тунис	-0.5443	0.1522	-0.5427	0.0541	-1.6049	0.1085
33	const	11.1326	0.6619	9.8352	12.4300	16.8180	0.0000
34	ln_height	-0.0545	0.1120	-0.2740	0.1650	-0.4869	0.6264
35	ln_width	0.1295	0.1373	-0.1396	0.3985	0.9432	0.3456
36	ln_depth	0.1272	0.0653	-0.0008	0.2551	1.9479	0.0514

## Нестабильные эффекты

### – Чемоданные модели:

- \* Исходная модель: +0.57\*\* для  $q=0.9$
- \* Бутстрэп: ДИ [-0.33, 1.46] включает 0  $\rightarrow$  эффект нестабилен

### – Высота изделия (ln\_height):

- \* Исходно значима для  $q=0.1$  ( $p = 0.005$ )
- \* После бутстрэпа  $p = 0.248 \rightarrow$  эффект исчезает

### – Натуральная кожа:

- \* Исходно сильный эффект для  $q=0.9$  ( $p < 0.001$ )
- \* После бутстрэпа  $p = 0.144 \rightarrow$  не подтверждается для премиум-сегмента

Бутстрэп-анализ выявил существенную нестабильность эффектов для части переменных. Ключевые показатели цены (итальянское производство, премиальные бренды) подтвердили свою устойчивость а дизайнерские и цветовые факторы требуют дополнительного исследования

## 8.3 Тест Вальда

Проведем тест Вальда для всех переменных на значимое различие коэффициентов для разных квантилей

$$H_0 : [q = 0.1]param.coef = [q = 0.9]param.coef$$

для каждого параметра

	Переменная	Коэф. (q=0)	Ст. ошибка (q=0)	Коэф. (q=0.5)	Ст. ошибка (q=0.5)	Коэф. (q=0.9)	Ст. ошибка (q=0.9)	Р-значение	Wald stat	P-value	One-tailed P
28	manufacture_country_Италия	0.539116	0.126544	1.114474+00	0.122205	-0.575359	0.175961	-3.269810	0.001076	Да	
7	brand_Borbonese	-0.477181	0.119147	-1.100724+00	0.237326	0.648911	0.265556	2.443598	0.014542	Да	
13	brand_MICHAEL Michael Kors	-0.700602	0.250304	-2.345324+00	0.500484	1.479540	0.819399	3.888244	0.014902	Да	
9	ln_height	-0.326619	0.255578	-7.878715+01	0.405239	-1.145430	0.522075	-1.344609	0.032777	Да	
25	design_country_Соединенное Королевство	0.547305	0.288467	-1.589394+01	0.178497	0.708296	0.338157	2.088669	0.038738	Да	
33	color	8.757724	0.417713	1.027161+01	0.681926	-1.551886	0.782707	-1.936722	0.052779	Нет	
12	brand_MAC JACOBI (NY)	-0.180357	0.317836	-1.462944+00	0.615897	1.328387	0.702293	1.868851	0.038939	Нет	
17	color_Коричневый	0.209437	0.087397	-1.033326+02	0.111432	0.219770	0.149156	1.551872	0.126993	Нет	
24	design_country_США	0.590052	0.448448	1.502581+00	0.557301	-1.002529	0.706633	-1.418741	0.155974	Нет	
35	ln_width	0.227813	0.135107	-2.430894+02	0.137281	0.251282	0.190399	1.308842	0.190588	Нет	
10	brand_Dale & Dalarna	0.887116	0.114482	3.116504+01	0.447508	0.371465	0.474722	1.207395	0.227396	Нет	
23	design_country_Франция	0.320116	0.289915	-5.994984+02	0.123832	0.380111	0.313259	1.205711	0.227629	Нет	
3	dual_cover	0.067337	0.158255	4.640735+01	0.325568	-0.387337	0.361886	-1.097657	0.272354	Нет	
8	brand_Bottega Veneta	1.106080	0.122688	6.470614+01	0.442002	0.459874	0.459282	0.999395	0.317604	Нет	
18	color_Красный	0.100720	0.189169	-1.151604+01	0.151491	0.235986	0.249440	0.940559	0.344233	Нет	
26	design_country_Япония	0.634225	0.333394	1.883424+01	0.149381	0.325700	0.365795	0.880390	0.373257	Нет	
19	color_Светло-коричневый	-0.038073	0.128347	1.415484+01	0.179529	-0.179623	0.220689	-0.813917	0.415682	Нет	
15	color_Белый	0.123222	0.067013	5.425586+02	0.083746	0.089968	0.090490	0.745667	0.453869	Нет	
27	manufacture_country_Бразилия	-0.883070	0.133477	-3.895928+01	0.230702	0.277522	0.405503	0.684596	0.493725	Нет	
32	manufacture_country_Тунис	-0.108800	0.179029	-2.623024+01	0.152235	0.151203	0.235004	0.651917	0.514465	Нет	
36	ln_depth	0.054860	0.067249	1.133534+01	0.065291	-0.058495	0.093730	-0.624081	0.532574	Нет	
29	manufacture_country_Казахстан	0.078454	0.196715	-1.198133+01	0.267964	0.198248	0.332417	0.596382	0.550020	Нет	
4	material_Кожа натуральная 100%	1.224693	0.560653	1.691636+00	0.523140	-2.404937	0.768517	-0.554517	0.579471	Нет	
21	material_Кожа натуральная 100%	0.259874	0.077728	8.926244+02	0.421071	0.254948	0.428185	0.501597	0.615670	Нет	
1	keychain	0.028816	0.222574	2.275200+01	0.348482	-0.198710	0.418963	-0.474289	0.635204	Нет	
30	manufacture_country_Китай (Китайская Народная Республика)	-0.960358	0.131884	-1.302107+01	0.121844	0.009852	0.179407	0.389352	0.697016	Нет	
16	color_Фиолетовый	0.157262	0.157420	1.114462+02	0.425718	0.412036	0.457565	0.334652	0.738105	Нет	
20	color_Черный	0.024245	0.059175	8.264684+00	0.067349	0.024245	0.077736	0.311893	0.755122	Нет	
6	makeup_bag	0.032769	0.131388	-1.900679+02	0.105130	0.051775	0.188208	0.307805	0.758231	Нет	
5	second_bath	0.267574	0.219124	1.600794+01	0.298943	0.157495	0.370651	0.290016	0.771804	Нет	
14	brand_Care-Beckend	-0.024089	0.237808	-5.176671+01	0.470262	0.151778	0.549636	0.262329	0.777591	Нет	
31	manufacture_country_Мадагаскар	-0.378124	0.238745	-5.935537+01	0.616814	0.125429	0.661406	0.189540	0.849591	Нет	
9	brand_Coccinelle	-0.680126	0.081927	-6.504344+01	0.133652	-0.029891	0.156764	-0.189401	0.848778	Нет	
2	locks	0.131287	0.229287	2.320474+01	0.781116	-0.151618	0.816389	-0.124473	0.900841	Нет	
34	ln_length	0.160149	0.141098	1.046134+01	0.111999	0.014995	0.162189	0.091838	0.620886	Нет	
11	brand_Furla	-0.396880	0.106664	-2.273307+01	0.406296	0.030351	0.430054	0.072253	0.942400	Нет	
22	material_Текстиль 100%	-0.200034	0.110875	-1.767177+01	0.426884	-0.023262	0.439906	-0.050581	0.959660	Нет	

Анализ теста Вальда на значимые различия коэффициентов между квантилями

Найдены статистически значимые различия для 5 переменных:

1. manufacture country Италия ( $p=0.001$ ): коэффициент увеличивается с +0.54 ( $q=0.1$ ) до +1.11 ( $q=0.75$ ) то есть влияние итальянского производства усиливается для более дорогих сумок

2. brand Borbonese ( $p=0.015$ ): эффект усиливается с -0.48 до -1.13 то есть наблюдается негативное влияние бренда на цену возрастает в премиум-сегменте

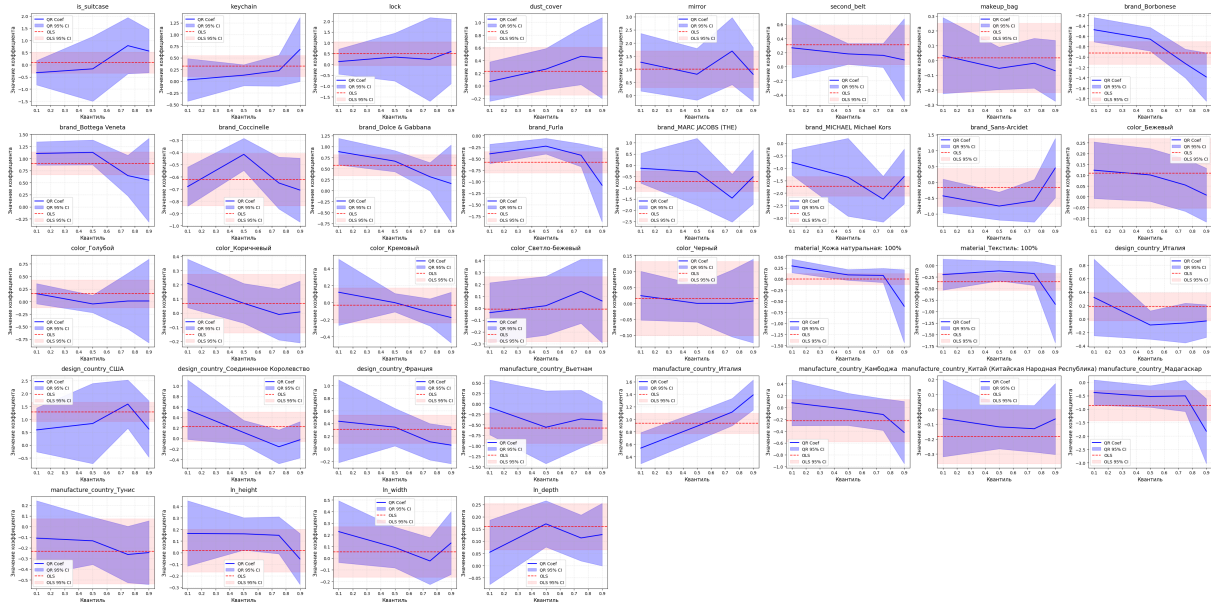
3. brand MICHAEL Michael Kors ( $p=0.017$ ): коэффициент падает с -0.77 до -2.25, отрицательное влияние усиливается в 3 раза для верхних квантилей



4. is\_suitcase (p=0.033): тут инверсия знака  $-0.33 \rightarrow +0.79$ , чемоданные модели снижают цену в массовом сегменте но повышают в премиумном

5. design\_country Соединенное Королевство (p=0.037): тут тоже положительный эффект (+0.55) сменяется отрицательным (-0.16)

## 8.4 Графики для каждого предиката



## Заключение

Работа проделана, мы молодцы