



Проектная работа по дисциплине  
"Эконометрика 2"

Факультет  
экономических наук

Москва, 2025

# Оценивание влияния различных факторов на цену одного грамма чая с использованием квантильной регрессии

Гаврин Глеб Юрьевич, БЭК2212  
Каленбет Ульяна Денисовна, БЭАД223  
Шалунова Анна Игоревна, БЭАД221



Москва, 2025

Проектная работа по  
дисциплине "Эконометрика 2"Оценивание влияния различных факторов на  
цену одного грамма чая с использованием  
квантильной регрессии

2

## Данные

Название	Цена за грамм (₽/г)	Описание	Регион	Ссылка	Аkkордеон: ЦВЕТ, АРОМАТ И ВКУС	Аkkордеон: СОСТАВ	Аkkордеон: КОЛИЧЕСТВО В УПАКОВКЕ	Аkkордеон: ВЕС ПАКЕТИКА (Г)
АЛТАЙСКИ	4.652	Авторская смесь	Россия	<a href="https://teaco.ru">https://teaco.ru</a>	ВкусНезначительная терпкость, умеренная кислинка, легкий цветочный аромат	шиповник, Облепиха целая сублимированная, зизифора		
АЛТАЙСКИ	9.94	Авторский боди	Россия	<a href="https://teaco.ru">https://teaco.ru</a>	ВкусЗначительная терпкость, умеренная кислинка, легкий цветочный аромат	иван-чай черный, сушеная клюква, лист брусники, саган-дайля, ягоды брусники, зизифора, плоды малины		
АРОМАТИЧЕСКИЙ	5.928	Классический	Китай	<a href="https://teaco.ru">https://teaco.ru</a>	ВкусСредняя терпкость, умеренная кислинка, легкий цветочный аромат	чай зеленый жасминовый		
ACCAM №1	2.728	Индийский черный	Индия	<a href="https://teaco.ru">https://teaco.ru</a>	ВкусЗначительная терпкость, умеренная кислинка, легкий цветочный аромат	чай черный		



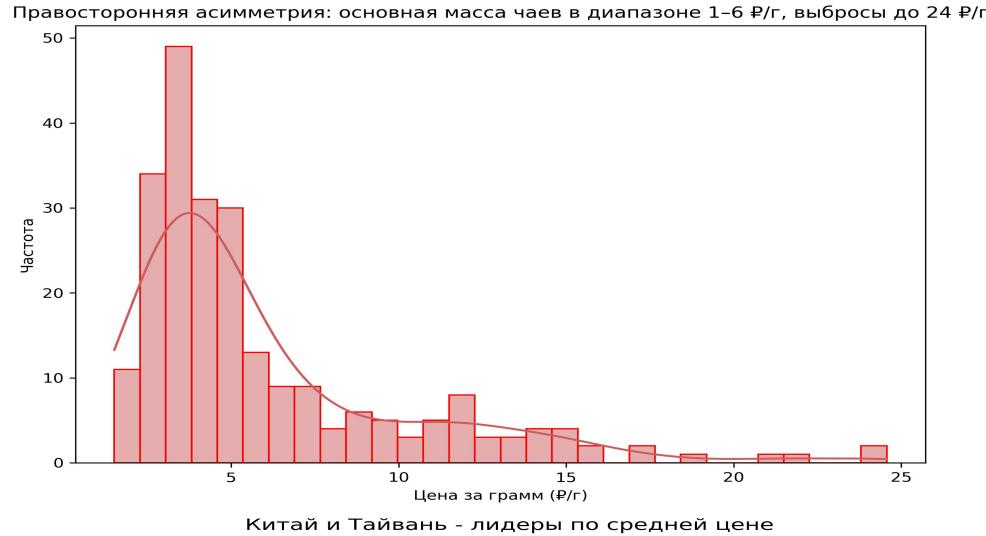
Название	Цена за грамм (₽/г)	Количество ингредиентов	Россия	Китай	Индия	Тайвань	Шри-Ланка	ЮАР	Купаж	Содержит ароматизатор	Непакетированный	Прозрачный	Непрозрачный	Интенсивный	Неинтенсивный	Терпкость	Кислотность	Продолжительное послевкусие	Горечь	Сладкость	Интенсивность послевкусия	
АЛТАЙСКИ	4.652	3	1	0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	0	1	2	
АЛТАЙСКИ	9.94	8	1	0	0	0	0	0	1	0	1	1	1	0	0	1	3	1	1	1	0	6
АРОМАТИЧЕСКИЙ	5.928	1	0	1	0	0	0	0	1	0	1	1	0	0	0	1	2	1	0	0	0	5
ACCAM №1	2.728	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	3	1	1	0	0	2
ACCAM №2	2.426666666666666	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	3	1	1	0	0	2
ACCAM №3	7.396	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	3	1	1	0	0	2
ACCAM №4	3.944	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	3	2	1	0	0	2
ГЛИНТВЕЙН	4.172	12	1	0	0	0	0	0	0	1	1	1	1	0	1	0	3	1	1	0	0	2
ГРАФОРМЕН	4.44	6	1	0	0	0	0	0	0	0	1	1	1	0	1	0	2	1	1	0	0	2

Источник данных: <https://teaco.ru/>Количество объектов: 240Парсинг: Selenium

В процессе отказались от признаков: «Фрукты», «Цветы», «Травы», «Острота», «Орехи», «Кислота»



## Анализ распределения цены и ключевых категориальных переменных



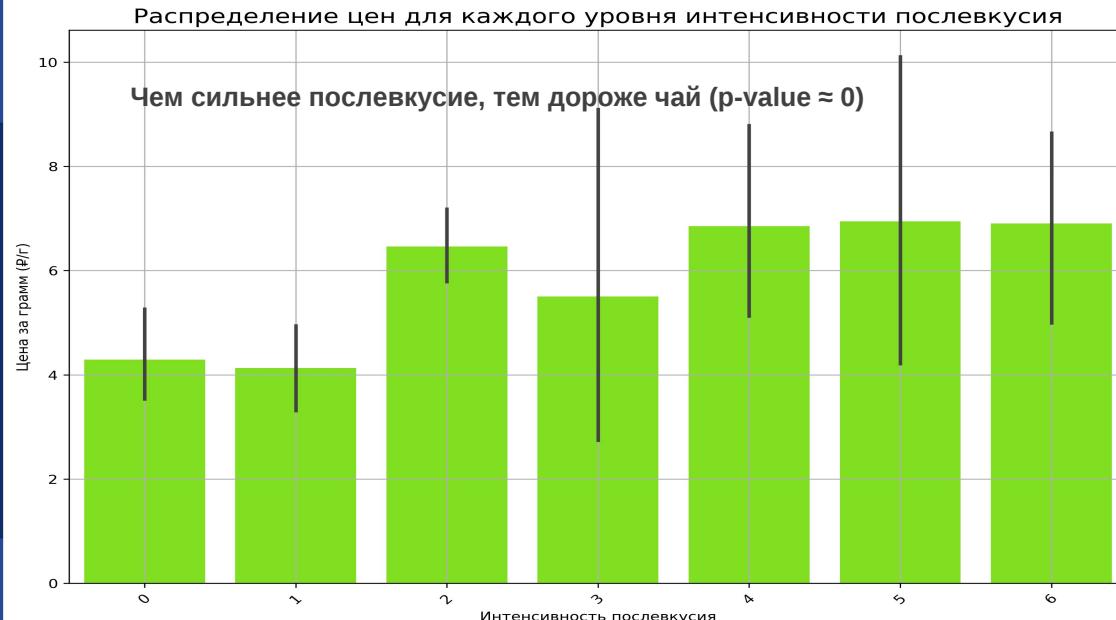
- Логарифмирование цены должно улучшить модель из-за асимметрии
- Моночай (1 ингредиент) дороже купажей ( $p\text{-value} \approx 0$ , тест Манна-Уитни)
- Китай и Тайвань - премиальные производители
- Россия - самый частый производитель, но цены на российский чай ниже

Средняя цена по выборке - **6.04 ₽/г**

Доля купажей - **40%**

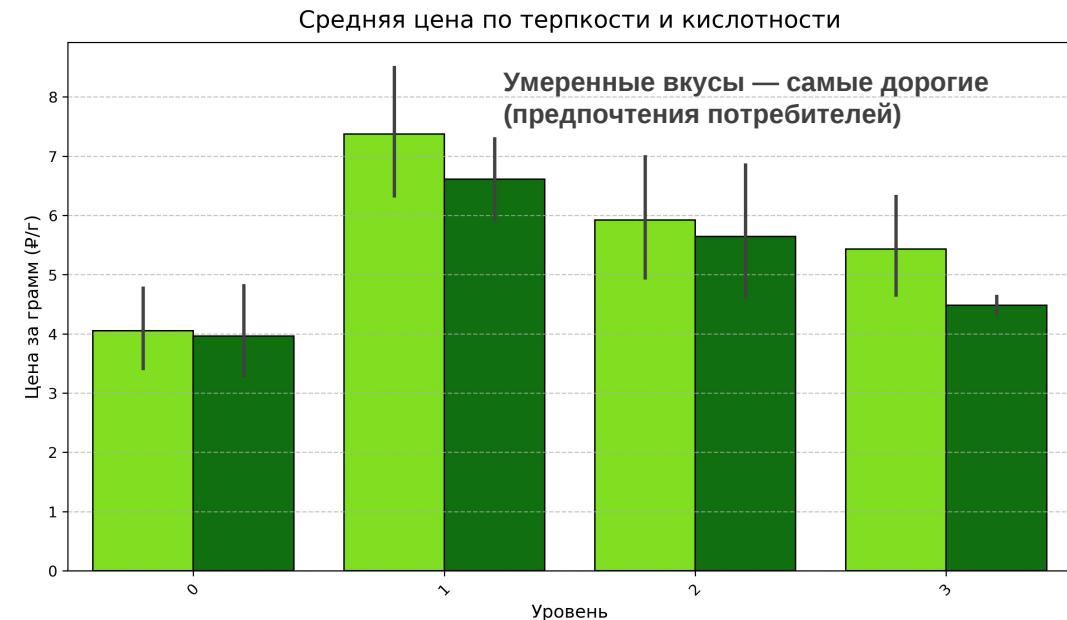


## Влияние вкусовых свойств на цену чая



### Результаты тестов:

- Послевкусие: группы 0-1 и 2-6 различаются ( $p\text{-value} \approx 0$ )
- Наличие ароматизатора: цены различаются ( $p\text{-value} \approx 0$ ), но средние близки
- Горечь: не влияет на цену (90% чаев без горечи)



### Выводы:

- Производители делают ставку на умеренную кислотность и терпкость - это максимизирует спрос
- Отсутствие горечи - стандарт для 90% рынка



## Итоговая модель

Цена за  $\widehat{\text{грамм}} (\text{₽}/\text{г}) = 1.709^{***} + 0.023 \cdot \text{Количество ингредиентов} - 0.179 \cdot \text{Россия}^* +$   
 $+ 0.322 \cdot \text{Китай}^{**} + 0.535 \cdot \text{Тайвань} - 0.120 \cdot \text{Купаж} -$   
 $- 0.174 \cdot \text{Содержит ароматизатор}^{**} - 0.394 \cdot \text{Непакетированный}^{***} +$   
 $+ 0.386 \cdot \text{Прозрачный}^{***} + 0.283 \cdot \text{Непрозрачный} - 0.231 \cdot \text{Интенсивный}^{***}$   
 $+ 0.176 \cdot \text{Продолжительное послевкусие}^{**}$

- 1) Была использована OLS с робастными стандартными ошибками (`cov_type='HC3'`).
- 2) Был произведен выбор между линейной, полулогарифмической и логарифмической функциональными формами с помощью теста Бокса-Кокса.

Данная модель прошла тест на адекватность:  $p\text{-value} \approx 0$ .

Также была попытка отобрать признаки по наименьшему AIC, но у получившейся модели вышел  $R^2\text{-adj}$  меньше, чем у итоговой.



## Анализ модели: гетероскедастичность и мультиколлинеарность

### Гетероскедастичность

Проблема была выявлена с помощью тестов Уайта, Глейзера, Бройша Пагана, а так же визуального анализа графика “остатки-прогнозы”

Решение: использовать OLS с робастными ошибками (НС3). Благодаря этому стали значимыми регрессоры “Россия” и “Наличие ароматизатора”

**Гетероскедастичность в модели локальная**, так как лишь для двух регрессоров значительно выросли стандартные ошибки после применения модели с робастными ошибками (“Китай” +21%, “Тайвань” + 60%)

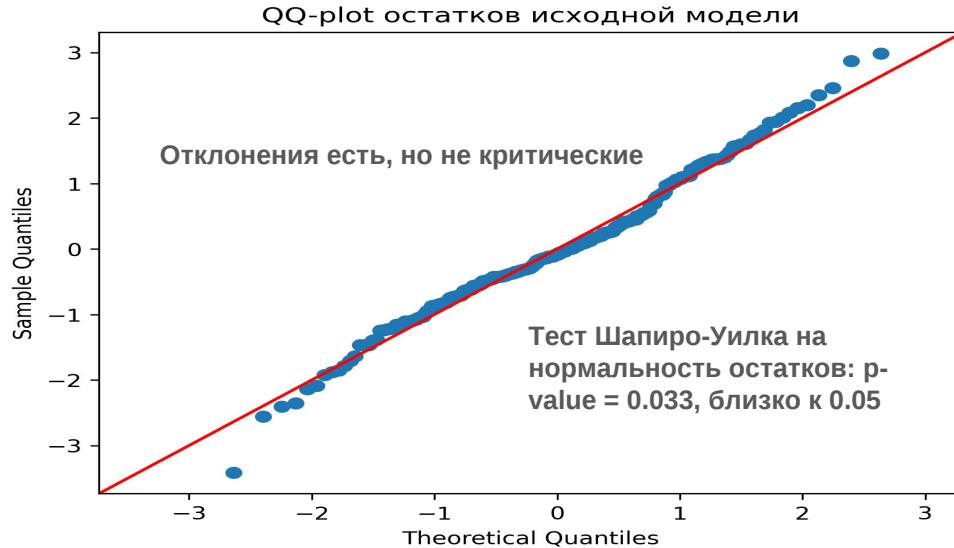
### Мультиколлинеарность

- Анализ корреляционной матрицы показал, что высоких корреляций в модели нет
- $VIF < 10$  для всех переменных

=> мультиколлинеарности в модели не обнаружено



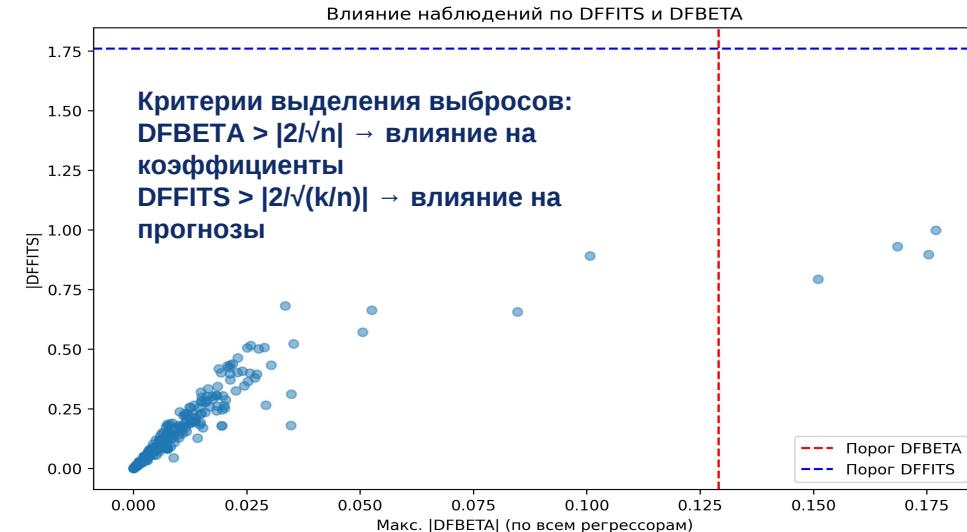
## Нормальность остатков и анализ выбросов



Распределение остатков **близко к нормальному**

### Проверка устойчивости (бутстррап):

- Разница ширины ДИ < 10%
  - Пересечение ДИ > 93% для всех коэффициентов
- => **оценки корректны**



### Результаты анализа:

- **4 наблюдения** влияют на коэффициент «Китай»
  - **Нет** точек, превышающих порог DFFITS => **нет искажения прогнозов**
  - Бутстррап подтвердил значимость регрессора «Китай» (ДИ не включает 0)
- => **Выбросы оставлены, влияние несущественно**



# Квантильная регрессия

## Дорогой сегмент ( $q=0.9$ )

Цена за  $\widehat{\text{грамм}} (\text{₽}/\text{г}) = 2.385 + 0.013 \cdot \text{Количество ингредиентов} - 0.217 \cdot \text{Россия} + 0.617 \cdot \text{Китай} +$   
 $+ 0.760 \cdot \text{Тайвань} - 0.105 \cdot \text{Купаж} - 0.347 \cdot \text{Содержит ароматизатор} -$   
 $- 0.517 \cdot \text{Непакетированный} + 0.370 \cdot \text{Прозрачный} + 0.422 \cdot \text{Непрозрачный}$   
 $- 0.135 \cdot \text{Интенсивный} + 0.277 \cdot \text{Продолжительное послевкусие}$

## Средний сегмент ( $q=0.5$ )

Цена за  $\widehat{\text{грамм}} (\text{₽}/\text{г}) = 1.530 + 0.020 \cdot \text{Количество ингредиентов} - 0.253 \cdot \text{Россия} + 0.338 \cdot \text{Китай} +$   
 $+ 0.655 \cdot \text{Тайвань} - 0.047 \cdot \text{Купаж} - 0.090 \cdot \text{Содержит ароматизатор} -$   
 $- 0.332 \cdot \text{Непакетированный} + 0.473 \cdot \text{Прозрачный} + 0.515 \cdot \text{Непрозрачный}$   
 $- 0.253 \cdot \text{Интенсивный} + 0.159 \cdot \text{Продолжительное послевкусие}$

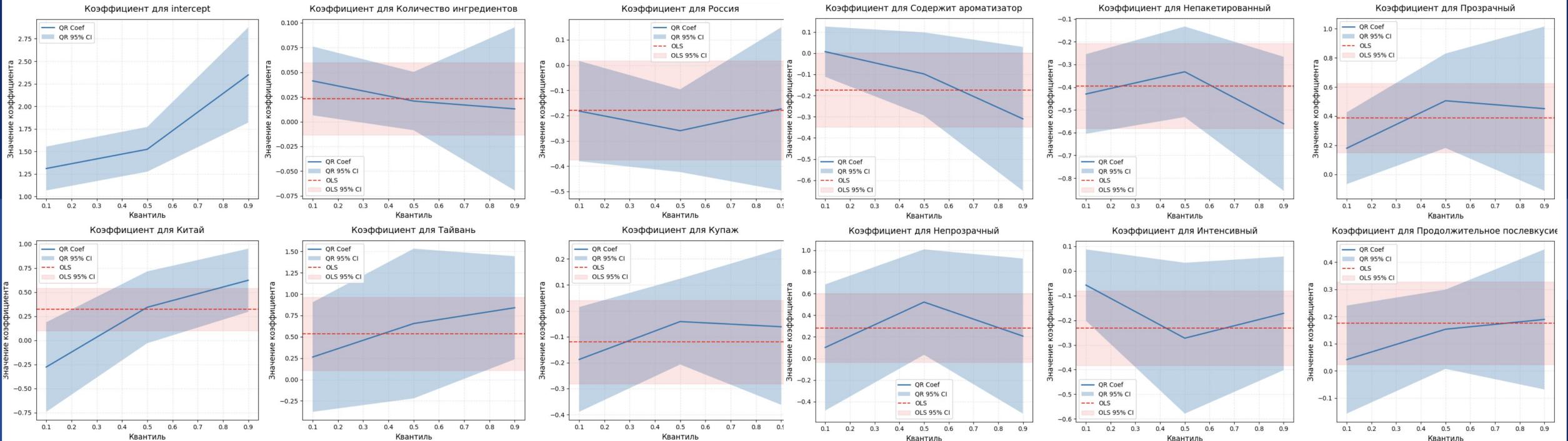
## Дешевый сегмент ( $q=0.1$ )

Цена за  $\widehat{\text{грамм}} (\text{₽}/\text{г}) = 1.310 + 0.041 \cdot \text{Количество ингредиентов} - 0.182 \cdot \text{Россия} - 0.277 \cdot \text{Китай} +$   
 $+ 0.263 \cdot \text{Тайвань} - 0.188 \cdot \text{Купаж} - 0.007 \cdot \text{Содержит ароматизатор} -$   
 $- 0.429 \cdot \text{Непакетированный} + 0.179 \cdot \text{Прозрачный} + 0.101 \cdot \text{Непрозрачный}$   
 $- 0.057 \cdot \text{Интенсивный} + 0.041 \cdot \text{Продолжительное послевкусие}$



## Квантильная регрессия: результаты

Тест Вальда выявил разницу в значимости для одного признака - «Китай», причем для всех ценовых сегментов.





## Выводы

- 1) Чай со средней и выше интенсивностью послевкусия имеют статистически значимую премию в цене по сравнению с чаями, у которых послевкусие слабое или отсутствует.
- 2) Однокомпонентные чаи в среднем значительно дороже купажей.
- 3) Чай из Китая в среднем значительно дороже, чем те чаи, у которых не указана страна производства. В то время как чаи из России дешевле.
- 4) Люди, покупающие чай в специализированных магазинах, таких как <https://teaco.ru/>, могут предпочитать более качественный чай — без ароматизаторов, неинтенсивный, но с продолжительным послевкусием.

