

# Отчет об исследовании набора данных о ценах на свеклу (2015-2019 гг.)

## Общая информация о наборе данных

- **Период данных:** 2015-2019 годы
- **Количество записей:** 18,250 строк
- **Основные столбцы:**
  - Дата
  - Средняя цена за единицу товара
  - Общее количество проданной свеклы
  - Количество малых свекл
  - Количество больших свекл
  - Количество очень больших свекл
  - Количество свекл, проданных в пакете

## Проверка статистических гипотез

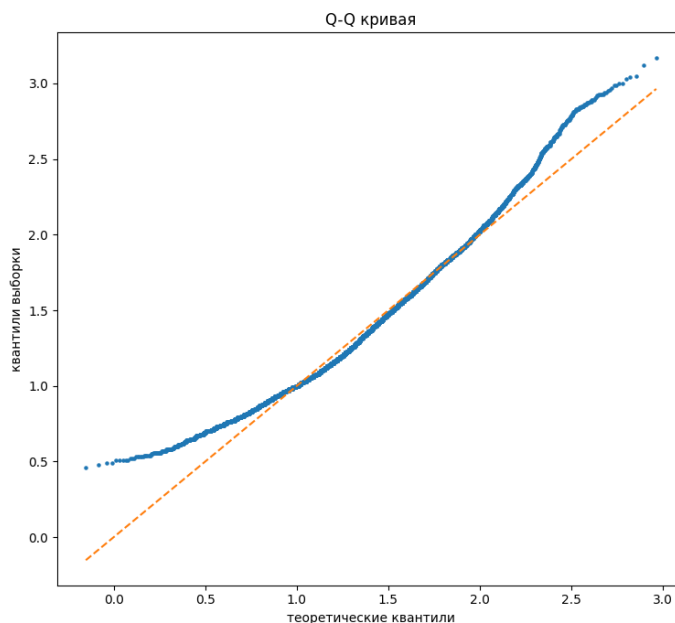
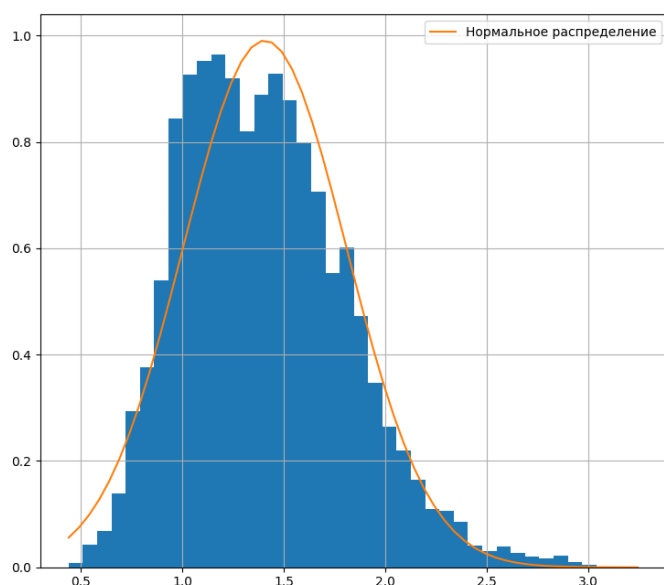
Гипотеза 1: Нормальность распределения средней цены

**Гипотеза:** Средняя цена распределена нормально.

### Методы проверки:

- Построение гистограммы распределения
- Q-Q plot (график квантиль-квантиль)

### Результаты:



**Вывод:** По визуальному анализу графиков наблюдается значительное отклонение от нормального распределения. Распределение средней цены не является нормальным или соответствует нормальному распределению с существенной погрешностью.

## Гипотеза 2: Наличие выбросов в данных

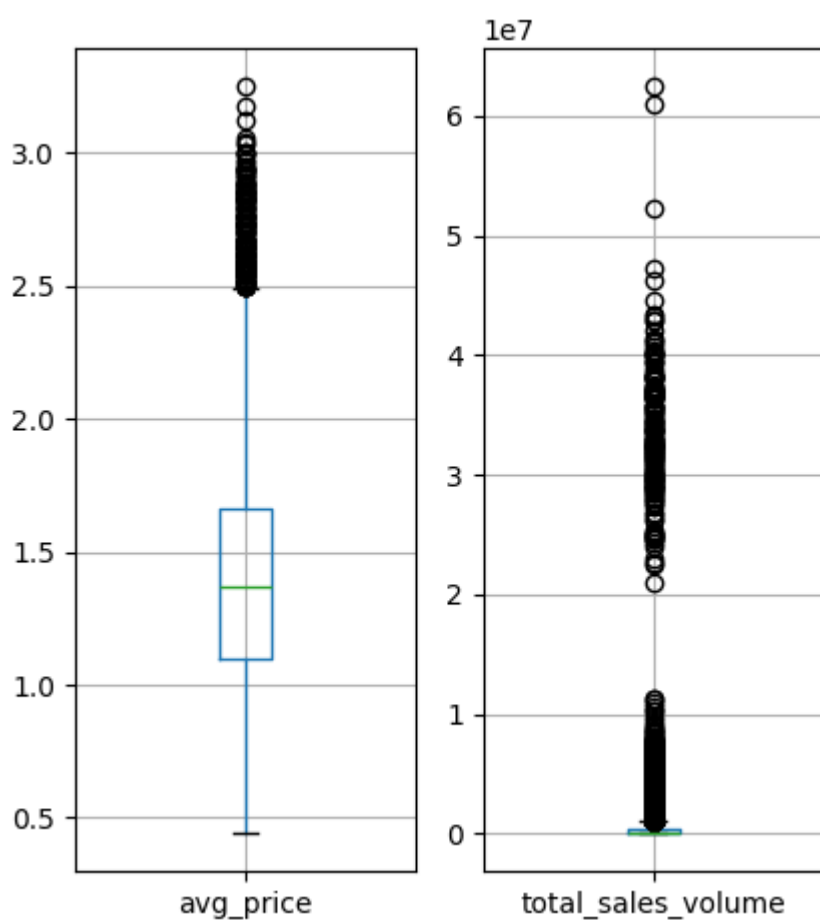
**Гипотеза:** В столбцах средних цен и общего числа продаж присутствует значительное количество выбросов.

### Методы проверки:

- Подсчет количества выбросов (по методу межквартильного размаха)
- Построение диаграмм "ящик с усами" (box plot)

### Результаты:

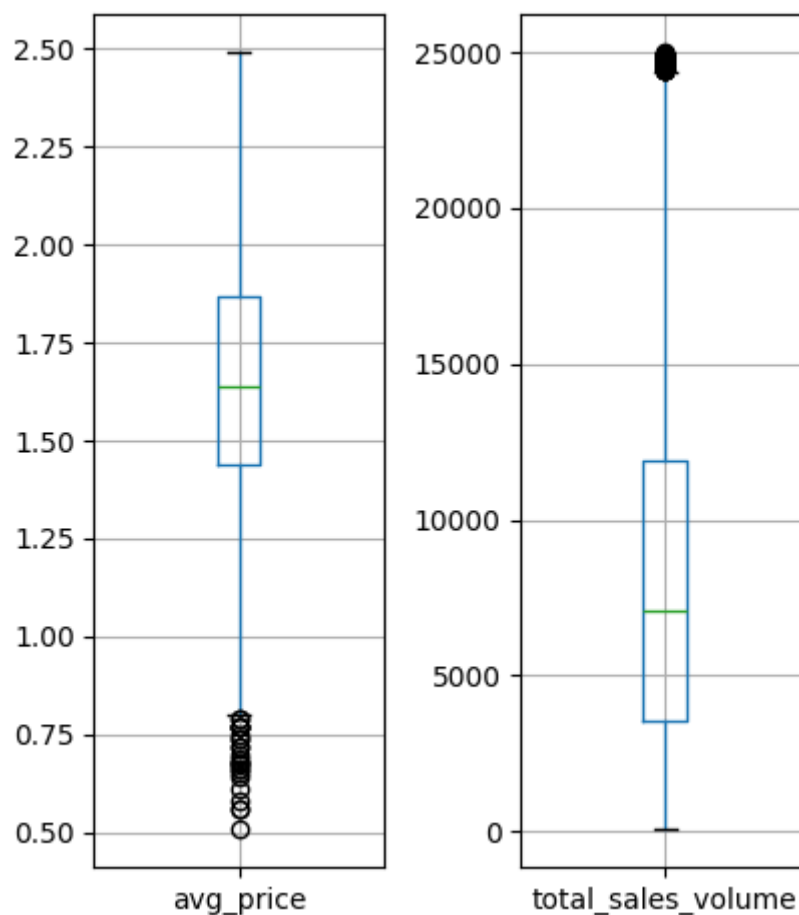
- Количество выбросов в столбце средней цены: **209**
- Количество выбросов в столбце общего числа продаж: **2297**



**Вывод:** Подтверждено наличие значительного количества выбросов в обоих столбцах, особенно в данных об общем количестве продаж.

## Обработка выбросов

**Действия:** Удалены записи с экстремально высокими значениями цен и объемов продаж.



**Результаты после обработки:**

**Наблюдение:** После удаления выбросов распределение данных стало более сбалансированным.

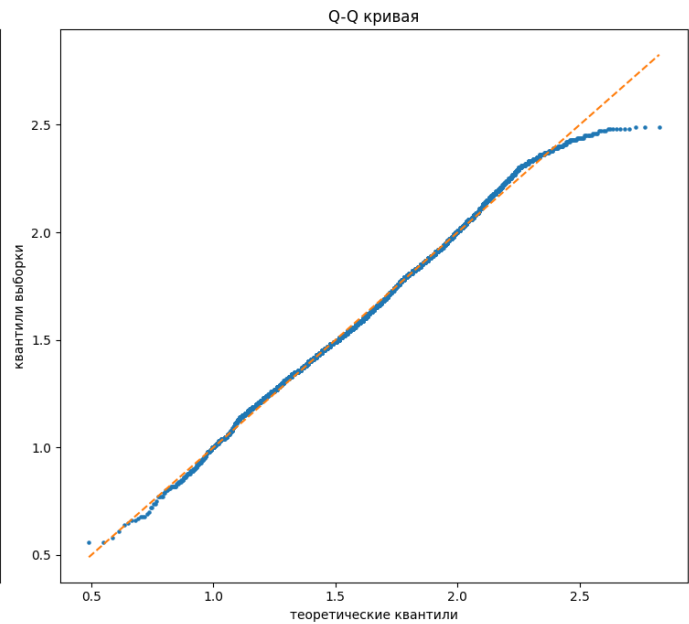
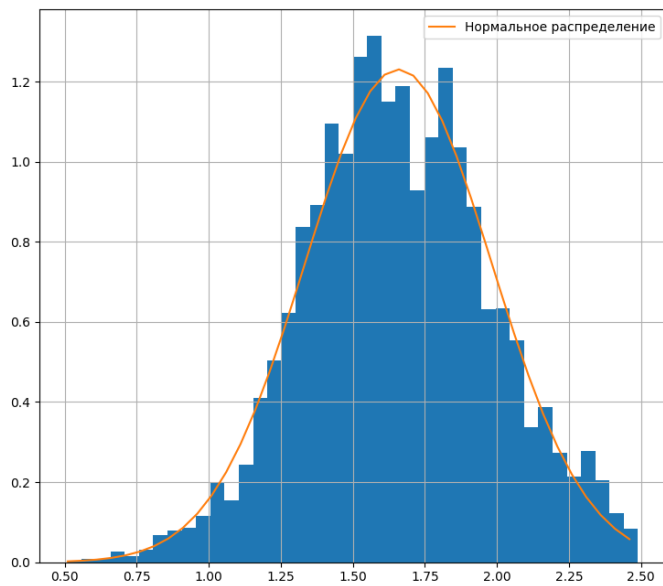
---

Гипотеза 3: Нормальность распределения цены после обработки выбросов

**Гипотеза:** После удаления выбросов цена распределена нормально.

**Методы проверки:**

- Построение гистограммы распределения
- Q-Q plot

**Результаты:**

**Вывод:** После удаления выбросов распределение средней цены хорошо соответствует нормальному распределению.

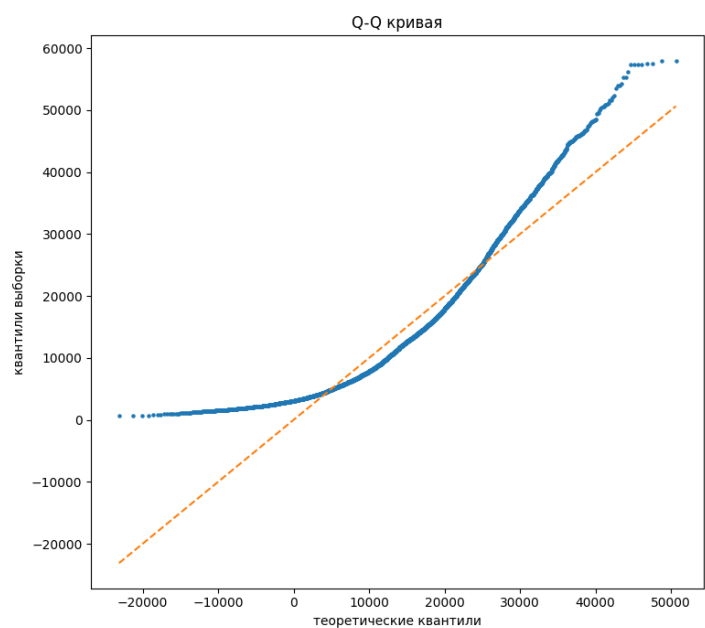
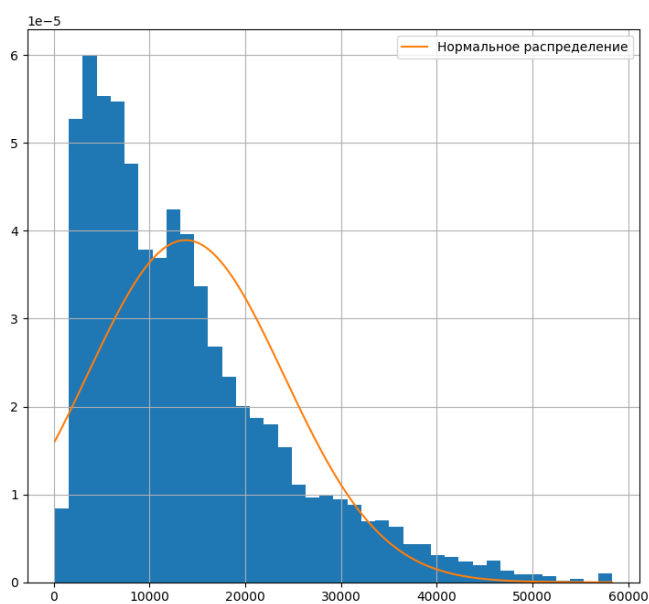
---

**Гипотеза 4: Нормальность распределения суммарной прибыли**

**Гипотеза:** Суммарная прибыль распределена нормально.

**Методы проверки:**

- Построение гистограммы распределения
- Q-Q plot

**Результаты:**

**Вывод:** Распределение суммарной прибыли совершенно не соответствует нормальному распределению.

---

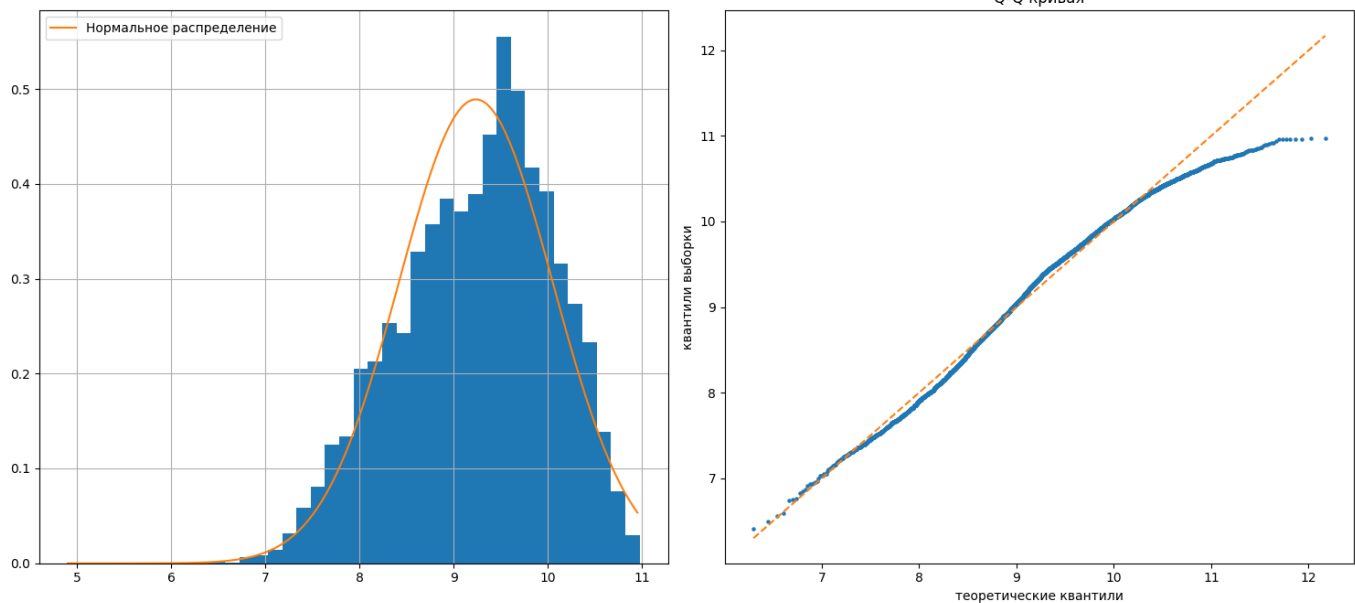
## Гипотеза 5: Логнормальность распределения суммарной прибыли

**Гипотеза:** Суммарная прибыль распределена логнормально.

### Методы проверки:

- Построение графиков для логарифма суммарной прибыли
- Гистограмма и Q-Q plot логарифмированных данных

### Результаты:



**Вывод:** Логарифм суммарной прибыли хорошо соответствует нормальному распределению, что подтверждает логнормальный характер распределения исходной переменной.

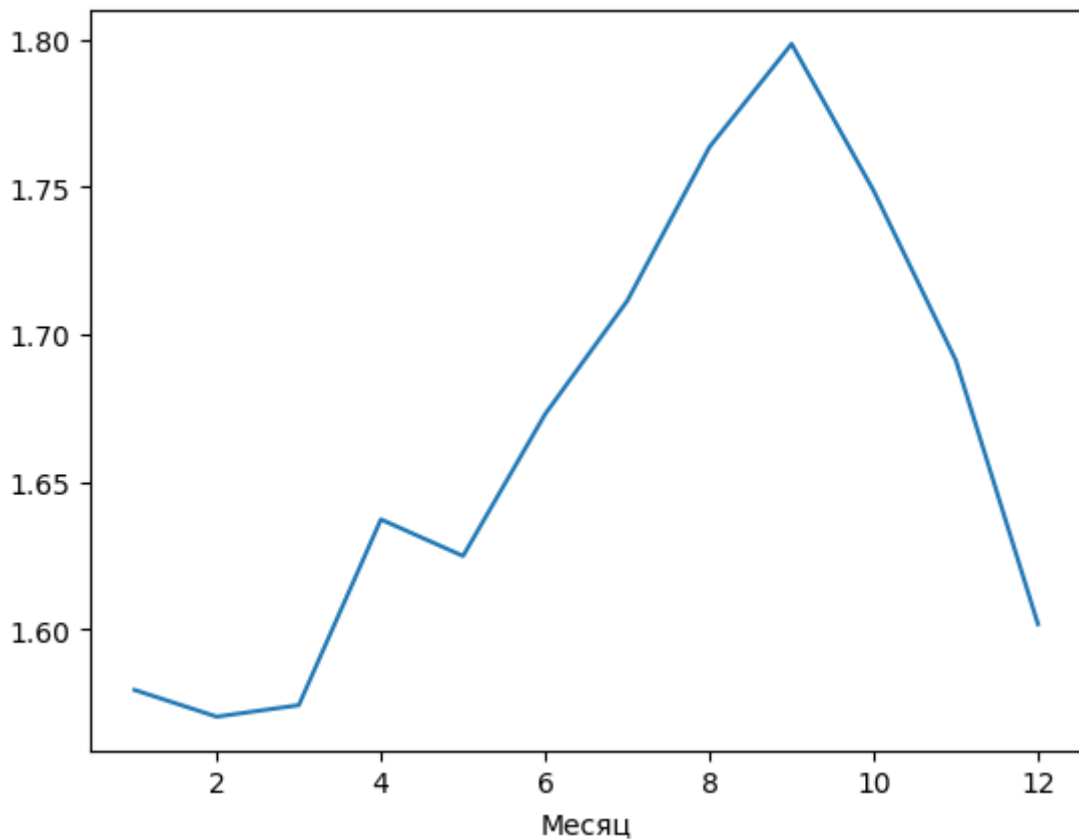
---

## Гипотеза 6: Сезонность цен

**Гипотеза:** Цена зависит от сезона.

### Методы проверки:

- Построение графика средней цены по месяцам года

**Результаты:**

- Наблюдается выраженная сезонность:
  - **Летние и осенние месяцы:** значительное повышение цен
  - **Зимние месяцы:** минимальные значения цен

**Вывод:** Подтверждена сильная сезонная зависимость цен на свеклу.

---

Гипотеза 7: Корреляция между ценой и объемом продаж

**Гипотеза:** Цена и общее число продаж коррелируют.

**Методы проверки:**

- Вычисление коэффициента корреляции

**Результаты:**

- Коэффициент корреляции: **-0.12**

**Вывод:** Обнаружена слабая отрицательная корреляция между ценой и объемом продаж. Связь между переменными статистически незначительная.