# Определение эффективности интенсивной технологии выращивания сахарной свеклы

|  |  |
| --- | --- |
|  | Профессия: Аналитик данных  Поток: DA-11  Труфанов Д.В. |

2021

Анализ проведен для сельскохозяйственного предприятия, отрасль – растениеводство.

Ежегодно в организации с сентября начинается процесс подготовки плана на следующий год. Основным этапом годового плана является определение севооборота и технологии возделывания сельскохозяйственных культур.

Технология – половина затрат компании. В организации более 550 тыс.га пашни, поэтому небольшое изменение стоимости технологии или увеличение урожайности может принести сотни миллионов прибыли или убытка.

В 2020 году на части полей была применена интенсивная технология выращивания сахарной свеклы, которая является более дорогой по сравнению по стандартной. Но данная технология позволяет получить с одного гектара больше готовой продукции. Фактические результаты получились неоднозначными.

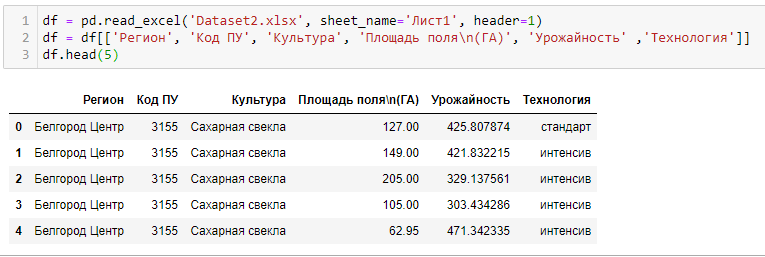
**Цель анализа:** определить эффективна ли интенсивная технология?

Для выполнения цели, решим ряд задач:

1. Изучим имеющиеся данные.
2. Проведем разделение данных.
3. Оценим урожайность по технологиям.
4. Оценим себестоимость.

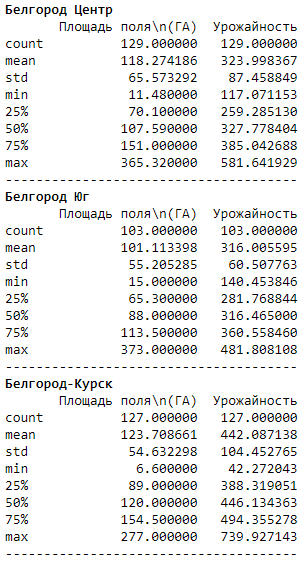
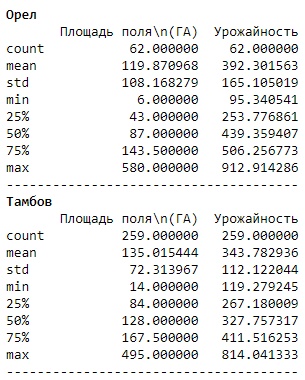
**Изучение данных**

Оставим в таблице только нужные поля:



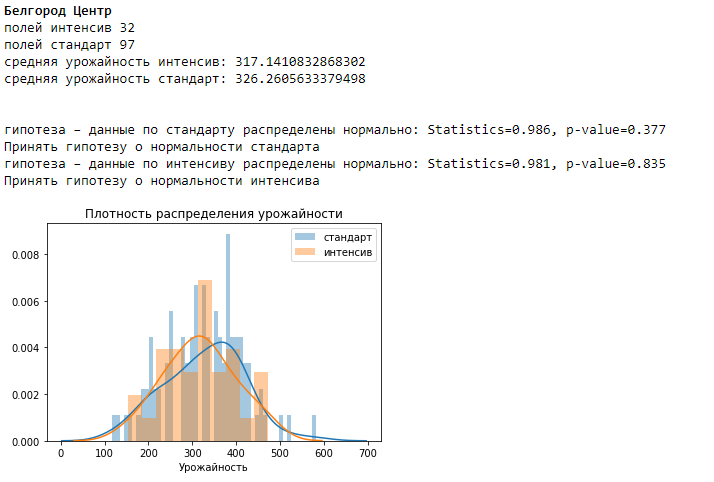
Сахарная свекла выращивается в регионах: Белгород Центр, Белгород Юг, Белгород-Курск, Орел, Тамбов.

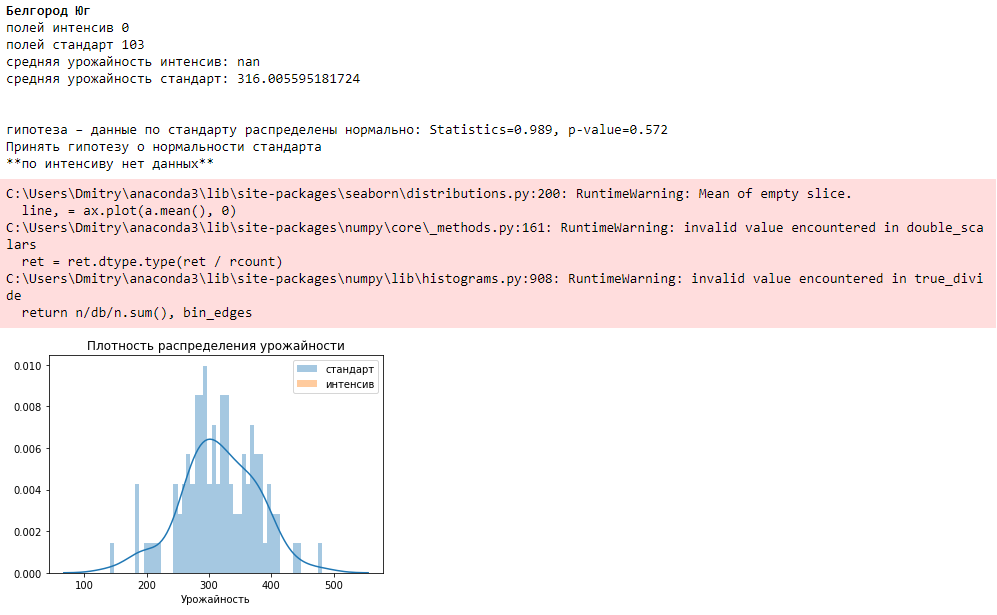
Мы знаем, что все столбцы содержат актуальную информацию, т.к. все значения фактические. Поэтому проверять набор данных на пропуски и выбросы не будем.

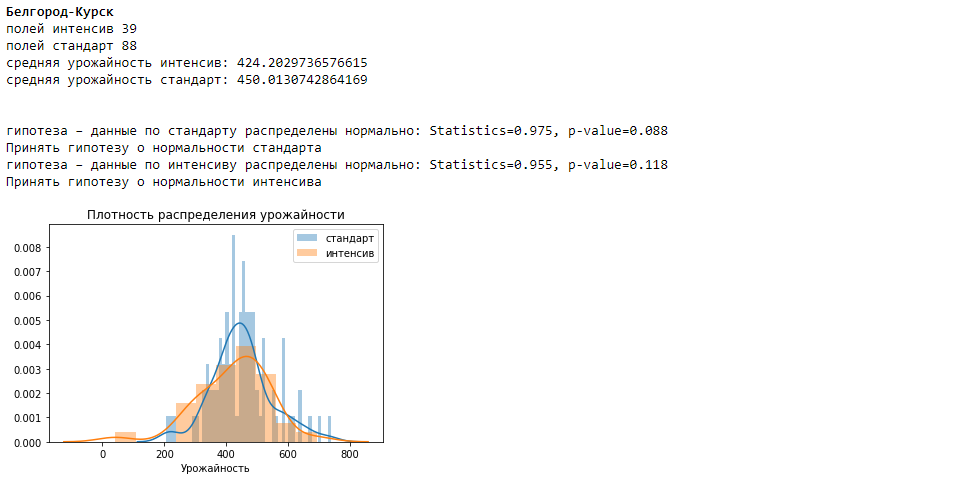
Проведем описание основных статистик для каждого региона.

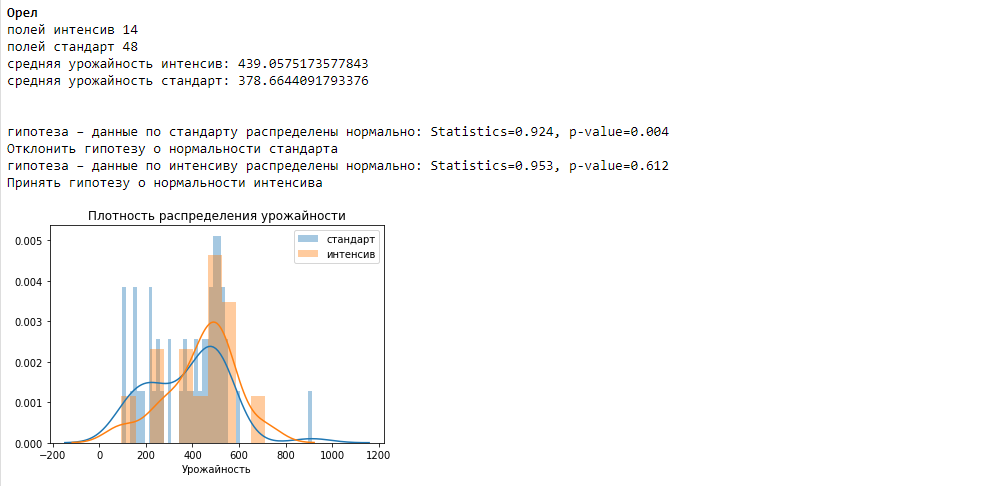
Выборки небольшие: от 62 до 259 полей. Урожайность неоднородна - минимальная и максимальная урожайности сильно отличаются.

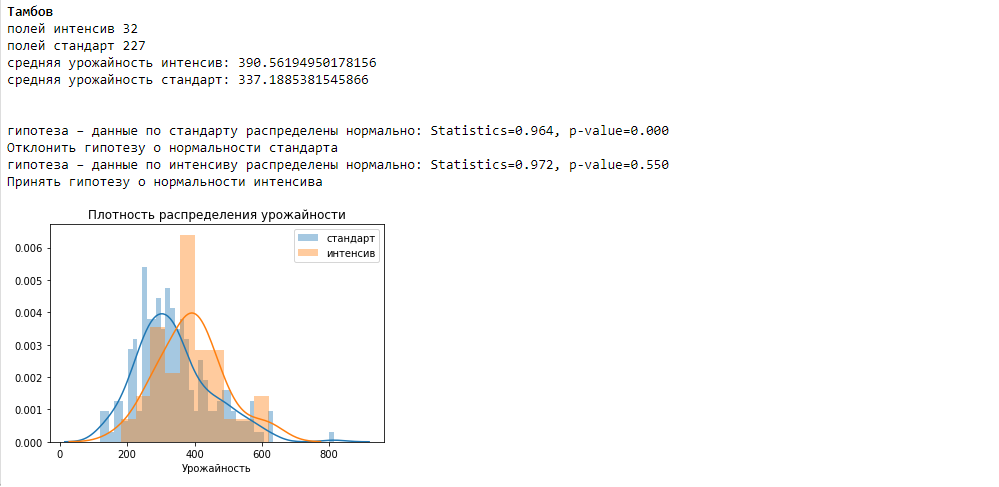
Для каждого региона определим размер выборки, среднюю урожайность по технологиям. Построим графики распределения урожайности для интенсивной и стандартной технологий, и при помощи теста Шапиро-Уилка проверим их на нормальность распределения.











Видим, что интенсив отсутствует в регионе Белгород-Юг и не во всех регионах урожайность по интенсивной технологии выше стандарта. В ряде регионов количество полей с интенсивной технологией очень мало. Для проверки гипотезы по различию урожайности, отбросим участки, с долей интенсива ниже 15%.

Для дальнейшего анализа опустимся с региона до уровня производственного отделения. Проведение анализа в рамках региона не целесообразно, т.к. поля имеют разные характеристики. Производственное отделение (ПУ) является наименьшим уровнем группировки, объединяющим небольшую группу полей схожими характеристиками.

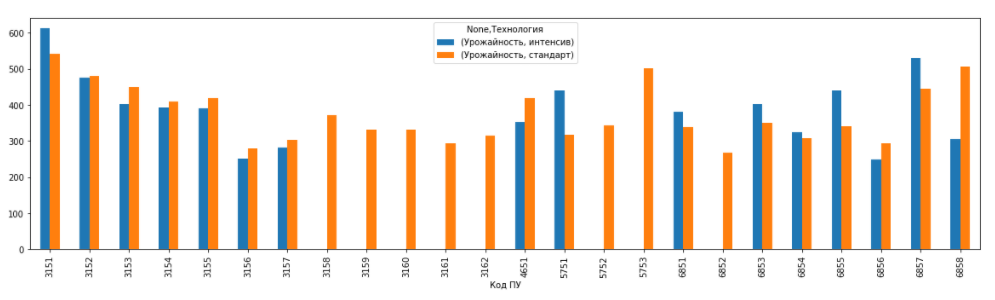


График показывает, что ситуация неоднозначна: в части ПУ интенсив более продуктивен, но есть ПУ с урожайностью по стандартной технологии, превышающей интенсив. Проведем построение модели по статистической проверке гипотезы отличия урожайности интенсива и стандарта.

**Проверка гипотез.**

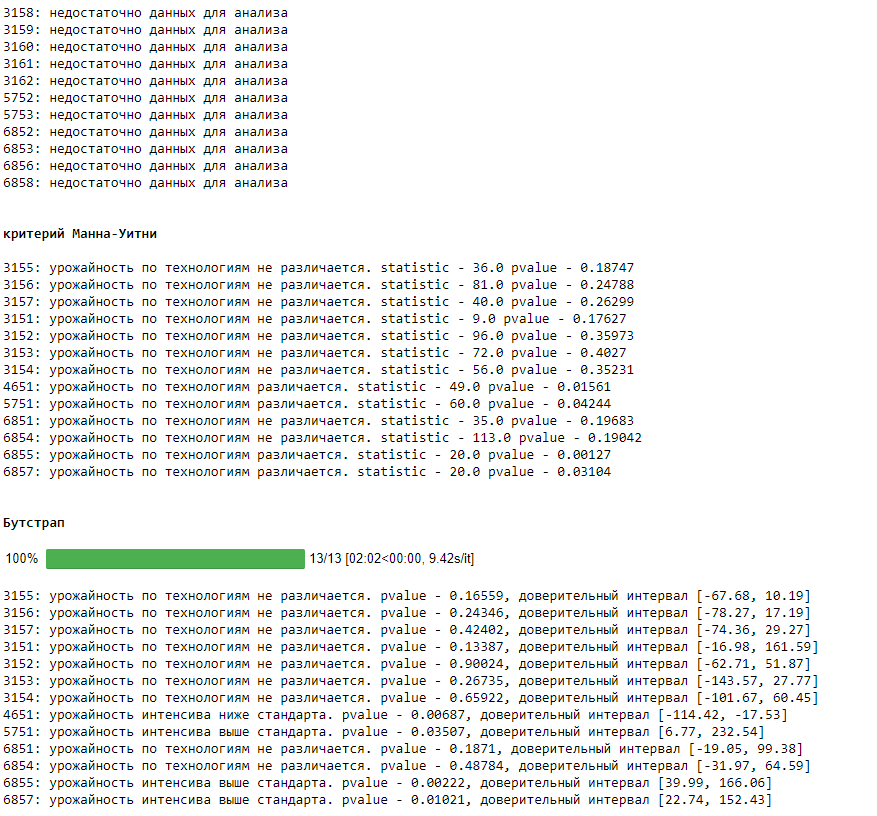
По части регионов распределение урожайности ненормальное, выборки небольшие. Поэтому для проверки гипотез будем использовать непараметрические статистические тесты.

Воспользуемся тестом Манна Уитни (ранжирование параметров из 2 выборок и подсчет суммы рангов для долей элементов первой и второй выборок) и Бутстрапом, основанном на многократной генерации выборок методом Монте-Карло на базе имеющейся выборки и определим доверительные интервалы.

**Результат расчета:**

Н0 – урожайности не отличаются

Н1 – урожайности отличаются



В 11 производственных отделения данных для анализа слишком мало.

По остальным ПУ оба теста показали схожие результаты:

В 9 производственных отделениях разницы в урожайности нет – **экономический эффект отрицательный**.

В 3 производственных отделениях урожайность при интенсивной технологии выше – **нужно оценить**.

В ПУ 4651 урожайность при интенсивной технологии ниже обычной, что может говорить о несоблюдении технологии – **экономический эффект отрицательный**.

Получается, проверку гипотезы об эффективности нужно проводить всего лишь для 3 участков.

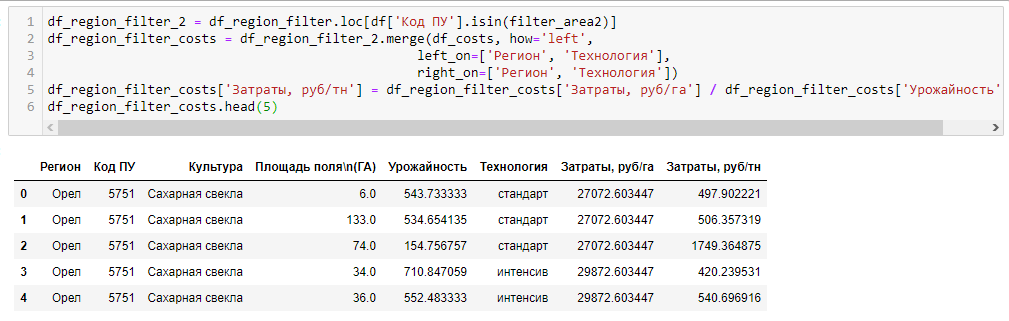
**Оценка себестоимости**

Загрузим датафрейм с затратами на га для каждого региона.



Объединим датафрейм по продуктивности полей с датафреймом по затратам.

Рассчитаем затраты на 1 тн.



На графиках посмотрим затраты на 1 тонну для разных технологий.

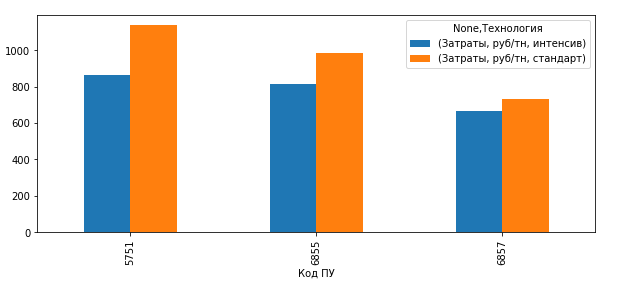
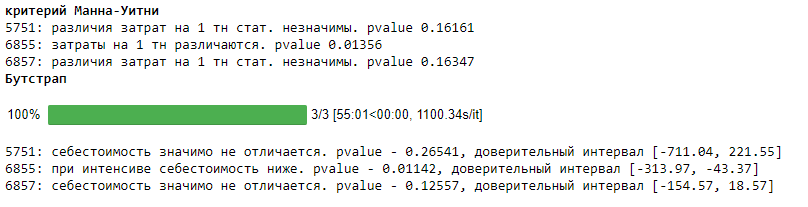


График показывает, что при интенсивной технологии затраты ниже.

Но, проверим гипотезу, на сколько разница в затратах статистически значима.



Оба теста снова дали одинаковый результат.

- в 2-х ПУ нет различий в себестоимости 1 тн. В таком случае использовать интенсивную технологию нецелесообразно, т.к. выше риск понести убытки.

**Вывод**

Таким образом, только в Тамбове ПУ 6855 интенсивная технология оправдала себя. Нужно с привлечением агрономической службой понять, каковы причины такого результата.

Оценим запас прочности эффективности интенсивной технологии:

**6855: на сколько можно изменить затраты на 1 га для равенства сс 1 тн по технологиям:**

фактические затраты на га: 34 801.0 руб.

максимально возможные затраты на га: 40 655.0 руб.

запас прочности на га: 5 854.0 руб.

**6855: изменение урожайности интенсива для уравнения себестоимости 1 тн при текущих затратах на га:**

фактическая урожайность интенсива: 444.0 ц/га

минимально возможная урожайность интенсива: 380.0 ц/га

запас прочности по урожайности: 64.0 ц/га.