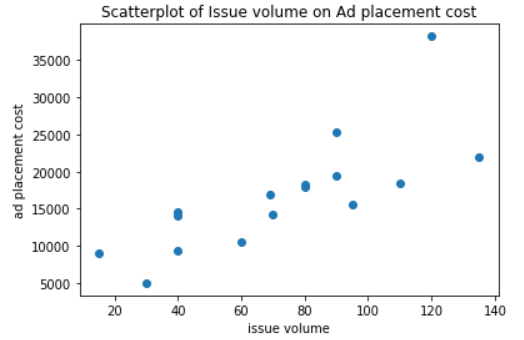
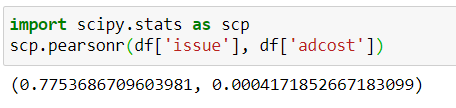
**1. Имеются данные о нескольких крупных изданиях и тарифы на размещение рекламы (цена модульного рекламного объявления)**



**а) постройте диаграмму рассеяния для зависимости тарифа на размещение рекламы от тиража газет**



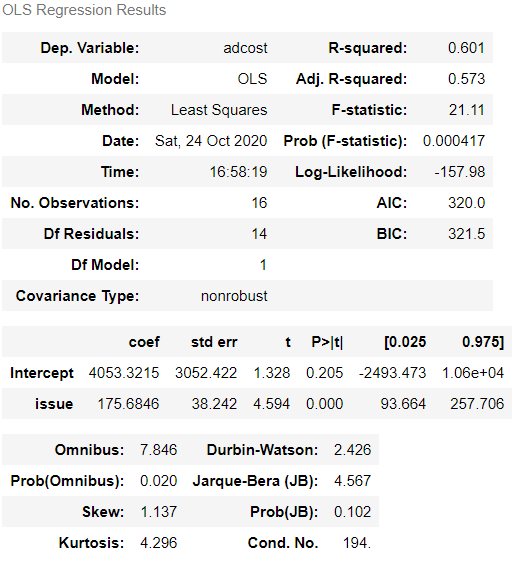
**б) дайте оценку корреляции тарифа на размещение рекламы и тиража. Можно ли объяснить эту корреляцию с экономической точки зрения?**



Первое значение – коэффициент корреляции r. При r = 0 – корреляции нет, при r = 1 – переменные линейно зависимы.

В данном случае r = 0.77 – это говорит о том, что связь между переменными есть, но не абсолютная. С экономической точки зрения это логично, поскольку чем больше тираж, тем дороже стоимость рекламы

**в) найдите уравнение регрессии, позволяющее прогнозировать тариф на размещение рекламы на основании тиража.**



Уравнение регрессии:

y = ax + b

intercept – это b (столбец coef)

issue – это a

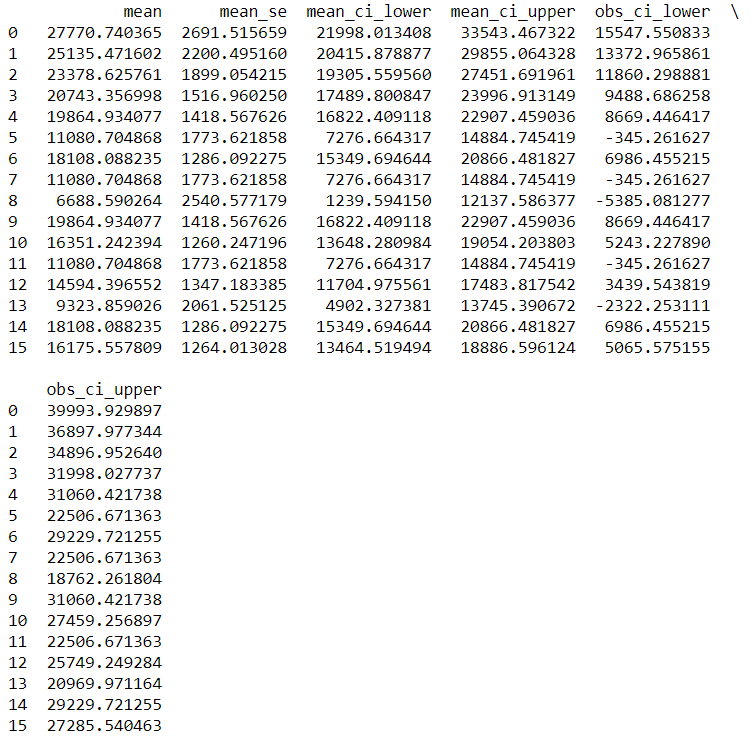
То есть:

Стоимость рекламы = Тираж (тыс.экз.) \* 175,68 + 4053,3 руб.

**г) проверьте, является ли статистически значимой связь между тарифом на размещение рекламы и тиражом;**

Поскольку P-значение = 0.000 (Это вероятность истинности нуль гипотезы, которая гласит, что независимые переменные не объясняют динамику зависимой переменной.), то отвергаем H0 (Она ниже, чем 0.05 - порог), то есть считаем, что связь статистически значима.

**д) найдите прогнозируемое значение и остаток для газеты Коммерсантъ. Ответьте на вопрос, является ли тариф на размещение одной строки рекламного объявления выше или ниже той величины, которую можно было бы ожидать от газеты с таким тиражом.**



Газета Коммерсантъ – в строке 0.

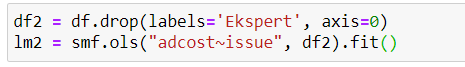
Прогнозируемое значение – mean – 27771

Остаток – mean\_se – 2692

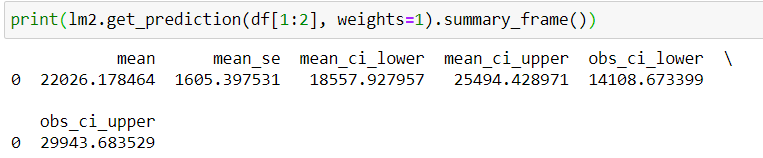
Тариф на размещение ниже той величины, которую можно ожидать, судя по построенной модели.

**2. По данным предыдущей задачи виден один возможный выброс (Эксперт). Необходимо выяснить, принадлежит ли это значение той же генеральной совокупности, что и остальные. С этой целью рассмотрим это значение как новое наблюдение.**

**а) Удалите Эксперт из рассматриваемой совокупности данных и найдите уравнение регрессии, позволяющее прогнозировать тариф на размещение рекламы на основании тиража для остальных газет;**



**б) Вычислите двусторонний 95% доверительный интервал для нового наблюдения, когда X0 является тарифом Эксперт;**



Доверительный интервал: X-1.96σ до X+1.96σ

Доверительный интервал: [18557,93 25494,43]

То есть с вероятностью 95% можно утверждать, что ожидаемая стоимость рекламы для журнала "Эксперт" должна составлять от 18557,93 руб. до 25494,43 руб., но фактическая стоимость – 38230.

**в) Проверьте, является ли Эксперт выбросом, выяснив, попадает ли тариф на размещение рекламы в этой газете в построенный доверительный интервал.**

Тариф – 38230 – не попадает в доверительный интервал

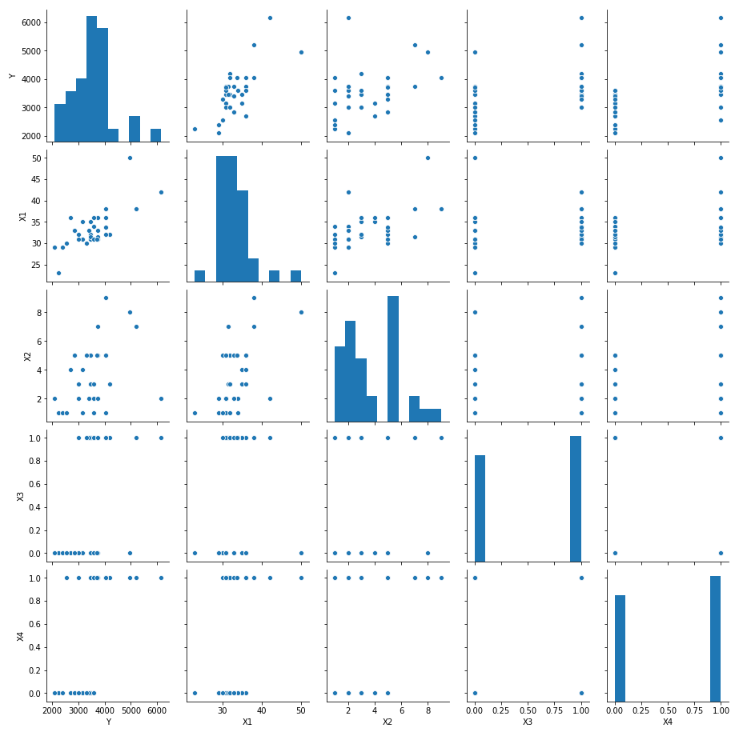
Следовательно, можно сделать вывод, что Эксперт – выброс

**В таблице представлены результаты исследования рынка жилья города. Необходимо построить аналитическую модель оценки стоимости жилья в зависимости от предложенных факторов.**

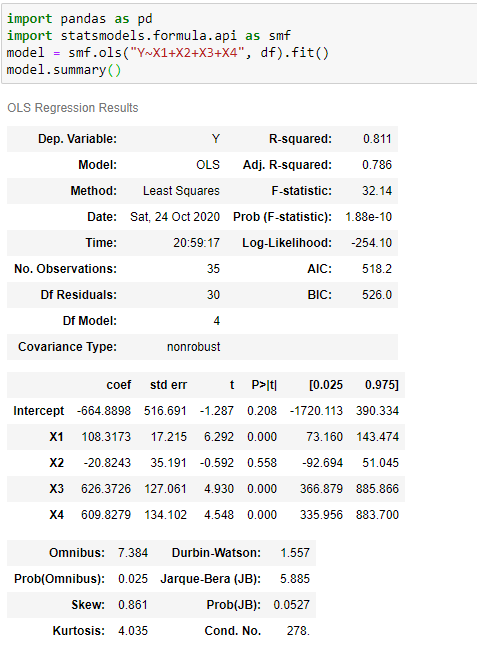
**Оцените параметры линейного уравнения множественной регрессии с фиктивными переменными.**

Построение попарных диаграммы рассеяния:

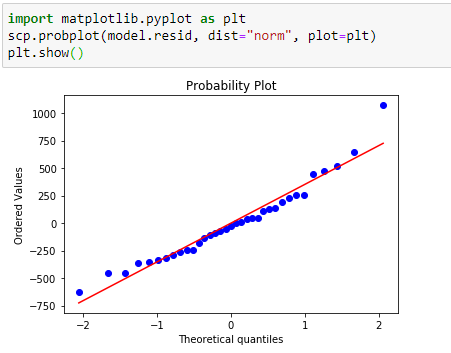




Построение модели:



Проверка распределения регрессионных остатков



Коэффициенты при параметрах говорят о том, что наибольшее влияние оказывают площадь и показатели наличия балкона и ремонта, что логично, поскольку чем больше площадь, тем выше стоимость и если есть ремонт/балкон, то стоимость также возрастает, а этаж показывает отрицательную зависимость, то есть чем выше этаж, тем ниже стоимость квартиры, при этом этот параметр мы не считаем статистически значимым, поскольку P-значение при нём превышает порог – 0.05