

**Зимняя школа “Эконометрика и анализ данных: программирование в R и Python”**  
**Домашняя работа №1 Заяц Софьи Александровны,**  
[sazayats\\_1@edu.hse.ru](mailto:sazayats_1@edu.hse.ru)

**Как я использовала бутстрэп?**

**Данные:** был выбран датасет на основе измерений скорости света Майкельсоном сделанные с 5 июня по 2 июля 1879 года, состоящий из 100 наблюдений<sup>1</sup>. Чтобы данные загрузились на любом компьютере, я скопировала их вручную из csv файла, так как объем выборки это позволял сделать без больших временных затрат и риска что-то упустить.

**Оцениваемые параметры и перспективы:** в этом задании я ограничилась только оценкой среднего значения и медианы, пока без нахождения корреляций, так как на данном этапе развития эта часть сложна для меня, но с получением опыта она станет возможной.

**Процесс работы:** был загружен пакет arch для использования функции бутстрэпа. Истинное распределение нам неизвестно, но если увеличить размер до нескольких тысяч, то результат будет близок к истинной функции распределения. Приступим к генерации выборок используя данные из исходной выборки: из наших  $n=100$  наблюдений случайно выберем 100 с возможностью повтора (генерируем выборки того же размера, что и исходная выборка, чтобы оценка стандартного отклонения была несмещённой). Сгенерируем 1000 выборок и вычислим интересующие нас статистики (этого будет достаточно для точного результата).

**ДЛЯ СРЕДНЕГО**

Среднее значение по изначальной выборке составляет 852.4, что является точечной оценкой. После генерации выборок и вычисления среднего для каждой из них, провизуализируем результат (рис. 1). Найдем 95% доверительный интервал для математического ожидания, убрав 2,5% самых маленьких и 2,5% самых больших значений. Получим результат: [836.2, 868.7025], в который вошло среднее по изначальной выборке.

---

<sup>1</sup> Источник: Michelson, A. A. (1882). "Experimental determination of the velocity of light made at the United States Naval Academy, Annapolis". Astronomical Papers, 1, 109-145, U. S. Nautical Almanac Office.

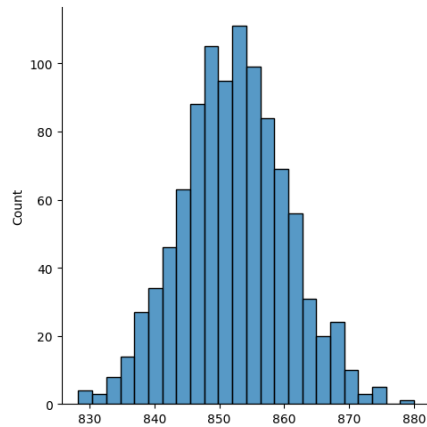


рисунок 1

### ДЛЯ МЕДИАНЫ

Проведем аналогичную процедуру, но для другой статистики — медианы. По изначальной выборке результат будет 850, а после применения бутстрэпа 95% доверительный интервал составит [840, 870] (см. рис. 2 с визуализацией для медианы).

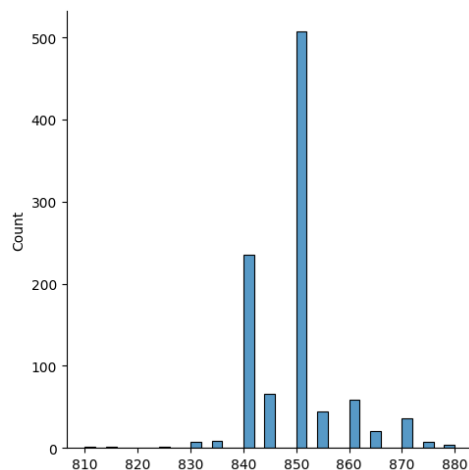


рисунок 2

**Результат:** в обоих случаях, полученные значения по исходной выборке попали в доверительный интервал после использования бутстрэпа. При попытке сгенерировать 10 тысяч выборок на основе исходной было замечено, что использование бутстрэпа имеет такой недостаток, как скорость работы. Тем не менее этот статистический метод позволяет получить интересующий нас результат без использования формул.

**Ссылка на код:**

[https://colab.research.google.com/drive/1RR2aCYHmt0k8A3sw1\\_Cx9njUo5x\\_u\\_7i?usp=sharing](https://colab.research.google.com/drive/1RR2aCYHmt0k8A3sw1_Cx9njUo5x_u_7i?usp=sharing)