

Для выполнения домашнего задания я решил взять часть тех данных, которые использовал в своей ВКР. Это информация об уровне конкурентоспособности экономик и об уровне защиты прав интеллектуальной собственности. Первый показатель был взят из отчетов Всемирного экономического форума (World Economic Forum), второй – с сайта Международного альянса прав собственности. Индекс конкурентоспособности представляет собой значения, которые могут находиться в диапазоне от 0 до 100 и принимают только целочисленные значения. Индекс защиты ПИС может принимать значения от 0 до 10. Данные взяты за 2019 год.

Сначала было посчитано среднее значение для показателя конкурентоспособности. Результат оказался следующим:

```
[10] np.mean(df.CI)

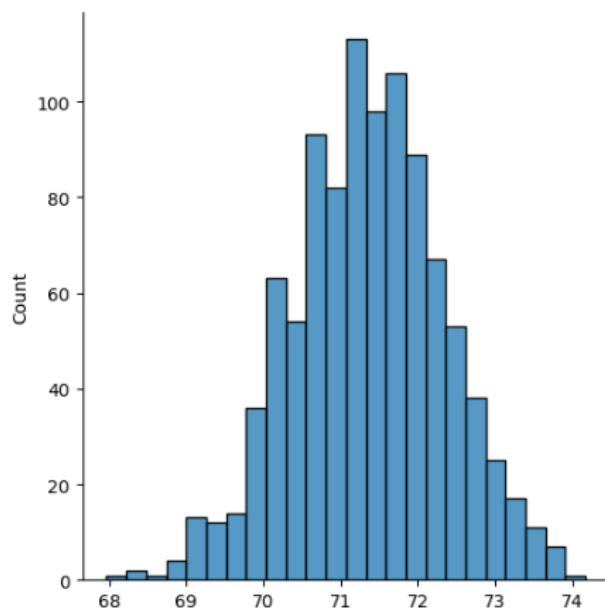
71.40350877192982
```

Затем при помощи бутстрэпа был посчитан доверительный интервал для математического ожидания:

```
[17] n_boot = 1000
mu_hat_star = [np.mean(rng.choice(df.CI, size=len(df.CI))) for i in range(n_boot)]
[np.quantile(mu_hat_star, 0.025), np.quantile(mu_hat_star, 0.975)]

[69.4030701754386, 73.2986842105263]
```

Из получившихся данных видно, что среднее значение лежит примерно посередине ДИ, а разница между ним и обоими концами интервалов не превышает 3%. Это может свидетельствовать в пользу гипотезы о нормальности распределения. Для визуального анализа построим график:



Из рисунка видно, что распределение данных похоже на нормальное.

Теперь рассмотрим показатель корреляции для наших данных. Вначале построим корреляционную матрицу:

```
np.corrcoef(df.CI, df.IRPI)

array([[1.          , 0.85700288],
       [0.85700288, 1.          ]])
```

Из результата можно заметить, что наблюдается сильная положительная корреляционная зависимость между показателями.

Теперь рассчитаем доверительный интервал при помощи бутстрэпа для истинной корреляции:

```
boot_xy = IIDBootstrap(x, y, seed=111111)
boot_xy.conf_int(corr, method='percentile', reps=10000, size=0.95)

array([[0.78484743],
       [0.89988002]])
```

Из полученного результата видно, что полученная изначально корреляция попадает в доверительный интервал, а её отклонение от его концов не превышает 8,5%, что может говорить в пользу нормальности распределения данных.