

1)

Имеется выборка из некоторого количества песен

Гипотезы:

H_0 : Популярность песен распределена нормально

H_1 : Популярность песен распределена НЕ нормально (как-то иначе)

Для проверки гипотезы проверим критерий Колмогорова-Смирнова и проведём тест Шапиро-Уилка(в коде 1.py)

Были получены следующие результаты:

Шапиро-Уилк:

p-value: ~0.0

Колмогоров-Смирнов:

p-value: ~0.0

Так как p-value оказалось равно нулю в обоих тестах, то на основе данных из предоставленной выборки скорее можно прийти к выводу, что выборка удовлетворяет гипотезе H_1 , а гипотеза H_0 отвергается.

2) Теперь сделаем разбиение нашей выборки на две подвыборки:

1) Песни до 4 минут

2) Песни больше 4 минут

Гипотезы:

H_0 : Обе подвыборки имеют одинаковое распределение

H_1 : Подвыборки имеют различные распределения

Полученные выборки проверим с помощью критериев однородности: теста Колмогорова–Смирнова и теста Уилкоксона-Манна-Уитни(в коде 2.py)

Тест Колмогорова–Смирнова:

p-value: 0.00004

Уилкоксона-Манна-Уитни:

p-value: 0.004

Так как p-value оказалось близким к нулю в обоих тестах, то на основе данных из полученных подвыборок скорее можно прийти к выводу, что выборка удовлетворяет гипотезе H_1 , а гипотеза H_0 отвергается.

3)

Проверим, зависит ли популярность песни от её длины. У нас есть две выборки: продолжительности песни и её популярность, проверим с помощью критериев независимости, есть ли корреляция между этими двумя параметрами.

Гипотезы:

H_0 : Популярность песен зависит от их продолжительности

H_1 : Популярность песен НЕ зависит от их продолжительности

Для этого используем коэффициент корреляции Кендалла и коэффициент корреляции Спирмена(в код 3.py)

Кендалл:

Коэффициент корреляции: 0.0119

p-value: 0.014

Спирман:

Коэффициент корреляции: 0.0180

p-value: 0.013

Так как p-value и коэффициенты корреляции оказались вновь близким к нулю в обоих тестах, то скорее можно прийти к выводу, что выборка удовлетворяет гипотезе H_1 , а гипотеза H_0 отвергается.