Практическая работа №5.

1. Для заданного набора данных (города России – town\_1959\_2.csv) проверить нормальность распределения по критерию Шапиро-Уилка. Предварительно построить гистограмму для исходной и для прологарифмированной совокупности (работа № )

res=stats.shpiro(np.log10(df[u’население’]))

print (‘p-value:,res[1])

Если p-value принимает значение меньше выбранного уровня значимости (а=0.05), гипотеза о нормальности отвергается. Проверить полученное утверждение можно визуально , увеличив число столбцов на гистограмме, и убедиться в наличии выбросов.

1. Для заданного набора данных (файл Albquerque Home Prices\_data.txt) проверить , что цены домов на углу(cor=1) и не на углу(cor=0) существенно отличаются.

Предварительно проверить нормальность данных(кр. Шапиро-Уилка),

проверить равенство дисперсий.

Сохранить две части выборки в отдельных переменных.

Если отличие от нормальности существенно - проверяется равенство медиан по критерию Манна-Уитни ( res=stats.manwhitney (….,…..))

(если нет- проверяется равенство средних по критерию Стьюдента).

Если p-значение достаточно велико, нет оснований отвергнуть гипотезу. Вывод , который нужно сделать в этом случае –цены домов различаются статистически незначимо-разница медиан носит случайный характер.

Примечания.

1. Не забыть подключить библиотеку stats.
2. Значение 9999 употребляется в качестве пустого значения, его необходимо заменить на корректное пустое значение NaN (replace(….,…))