

Практическая работа № 11

Проверка статистических гипотез. Построение модели регрессии.

Пример 11.1 Библиотеки для статистического анализа данных

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy
import scipy.stats as stats
import seaborn as sns
import sklearn.linear_model as lm
import statsmodels.api as sm
import statsmodels.formula.api as smf
```

Задание:

1. Подгрузить для анализа файл rent.csv
2. Структурируйте данные
3. Проанализируйте типы данных
4. Визуализируйте данные

Вариант 1 Вывести распределение количества объявлений от стоимости аренды.

Вариант 2 Вывести распределение стоимости в зависимости от района.

Вариант 3 Вывести распределение стоимости в зависимости от материала стен.

Вариант 4 Вывести распределение количества объявлений в зависимости от этажа.

Вариант 5 Вывести распределение количества объявлений в зависимости количества комнат.

5. Проверить критерий асимметрии и эксцесса для своего варианта распределения

Коэффициент асимметрии, у нормально распределенной выборки он равен 0 (возможны отклонения ± 1)

Коэффициент эксцесса, у нормально распределенной выборки он равен 3 (возможны отклонения ± 1)

```
fl[['rent', 's']].skew() # коэффициент асимметрии
fl[['rent', 's']].kurtosis() # коэффициент эксцесса
```

6. Проверить гипотезу о виде распределения для своего варианта распределения.

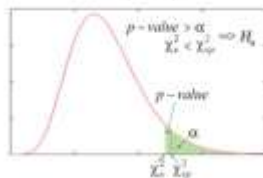
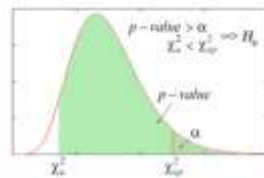
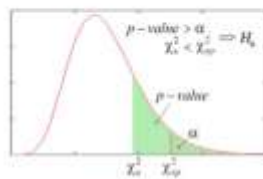
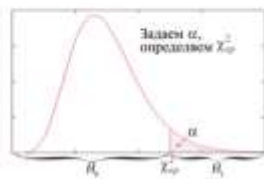
p-value - апостериорный уровень значимости

альфа - вероятность того, что получим значение статистики больше критической

p-value - вероятность того, что получим значение статистики больше чем на выборке

$$\alpha = P(H_1|H_0) = P(\chi^2 \geq \chi_{\alpha}^2 | H_0)$$

$$p\text{-value} = P(\chi^2 \geq \chi_{\alpha}^2 | H_0)$$



Тест д'Агостино

вычисляет итоговую статистику на основе данных, а именно, эксцесс и асимметрия, чтобы определить, отклоняется ли распределение данных от нормального распределения.

```
stats.normaltest(fl['rent'])
```

7. Оцените связь между показателями

В зависимости от типов данных, оцените связь между показателями.

Используйте критерии сравнения групп для анализа связи количественных или порядковых показателей с категориальными.

Выбор критерия сравнения групп



8. Сформулируйте свои гипотезы H0 H1. Проверьте их.

Например

Гипотеза H0: статистически значимые отличия в группах (тип стен дома) отсутствуют.

Гипотеза H1- статистически значимые отличия в группах (тип стен дома) наблюдаются.

Проверка по парному критерию Манна-Уитни

```
f1_0 = f1[f1['walls']==0]
f1_1 = f1[f1['walls']==1]
scipy.stats.mannwhitneyu(f1_0['rent'],f1_1['rent'])
```

9. Самостоятельно определите регрессионную модель. Для этого первоначально рекомендуется вывести диаграммы рассеяния.

Пример получения диаграммы рассеяния аренды от площади.

```
sns.jointplot('s', 'rent', data=f1, kind="reg")
```

Вывод статистики по регрессионному анализу :

```
sns.jointplot('s', 'rent', data=f1, kind="reg")
```