Необходимые библиотеки:

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import statsmodels.api as sm

import statsmodels.formula.api as smf

Необходимые формулы:



1. Создайте датафрейм, содержащий возможные исходы эксперимента по подбрасыванию монетки. Эксперимент проводился 10 раз. Вычислите ODDS.
2. Визуализируйте зависимость вероятности от логарифма шансов. Для этого вычислите

* вероятность успеха при каждом исходе,
* ODDS(шансы),
* натуральный логарифм ODDS.

Для работы используйте следующий датафрейм:

'pos': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10],

'neg': [10, 0, 8, 7, 6, 5, 4, 3, 2, 1, 0]

1. Постройте график зависимости вероятности от логарифма шансов.
2. Постройте модель (без предикторов)
3. Загрузите датасет: df = pd.read\_csv(‘train.csv')
4. Визуализируйте первые пять рядов:

df = df[df.Age.notnull()]

df.head()

1. Постройте логистическую модель без предиктора:

glm\_binomial = smf.glm(formula='Survived ~ 1', data=df, family=sm.families.Binomial()).fit()

glm\_binomial.summary()

1. Получите распределение частот (функция value\_counts())

Постройте count\_plot

(sns.countplot(x='Survived', data=df, palette='hls'))

1. Вычислите odds, ln(odds). Сравните полученные значения со значением Intersept.

Вычислите показатель остатков модели: glm\_binomial.**aic**