

### 3 Базовые статистические модели

**1. [# 30]** На станции лондонского метро подсчитывалось количество женщин в каждой из 100 очередей длиной 10. Таким образом, был получен набор данных  $x_1, x_2, \dots, x_{100}$ , где  $x_i$  обозначает наблюдаемое количество женщин в  $i$ -й очереди. Набор данных представлен в таблице

Женщин	0	1	2	3	4	5	6	7	8	9	10
Количество очередей	1	3	4	23	25	19	18	5	1	1	0

В статистической модели для этого набора данных мы предполагаем, что наблюдаемые значения являются реализацией случайной выборки  $X_1, X_2, \dots, X_{100}$ .

1) Предположим, что люди выстраиваются таким образом, что мужчина или женщина на определенной позиции независимы от других позиций, и что на каждой позиции с равной вероятностью находится женщина. Каков подходящий выбор для распределения модели? Постройте гистограмму и KDE для выдвижения гипотезы.

2). Найдите параметр(ы) модельного распределения, выбранного в части 1).

**2. [# 30]** Во время Второй мировой войны на Лондон обрушилось множество летающих бомб. Следующие данные относятся к площади в Южном Лондоне площадью 36 квадратных километров. Площадь была разделена на 576 квадратов со сторонами длиной  $\frac{1}{4}$  километра. Для каждого из 576 квадратов было записано количество ударов. Таким образом, мы получаем набор данных  $x_1, x_2, \dots, x_{576}$ , где  $x_i$  обозначает количество ударов в  $i$ -м квадрате. Данные обобщены в следующей таблице, в которой указано количество квадратов без ударов, с одним попаданием, с двумя попаданиями и т.д.

Попаданий	0	1	2	3	4	5	6	7
Количество квадратов	229	211	93	35	7	0	0	1

Вопрос заключается в том, велось ли бомбометание Лондона случайным образом? В этом случае данные должны иметь распределение Пуассона.

1). Если мы смоделируем набор данных как реализацию случайной выборки из распределения Пуассона с параметром  $\mu$ , то что бы вы выбрали в качестве оценки для  $\mu$ ?

2). Проверьте соответствие с распределением Пуассона, сравнив некоторые из наблюдаемых относительных частот 0, 1, 2 и т.д. с соответствующими вероятностями для распределения Пуассона с  $\mu$ , оцененными, как в части 1).

3) Постройте на одном графике кривые плотности для распределения Пуассона с параметром  $\mu$  и KDE для представленных данных.

3. [# 30] Исследовалось количество менструальных циклов до наступления беременности, где количество циклов моделировалось геометрической случайной величиной. Исходные данные касались 100 курящих и 486 некурящих женщин. Для 7 курильщиц и 12 некурящих точное количество циклов до наступления беременности было неизвестно.

1) Рассмотрим набор данных  $x_1, x_2, \dots, x_{93}$ , соответствующий курящим женщинам, где  $x_i$  обозначает количество циклов для  $i$ -й курящей женщины. Данные представлены в таблицах.

Циклы	1	2	3	4	5	6	7	8	9	10	11	12
Кол-во	29	16	17	4	3	9	4	5	1	1	1	3

В таблице указано количество женщин, которым пришлось ждать 1 цикл, 2 цикла и т.д. Если мы смоделируем набор данных как реализацию случайной выборки из геометрического распределения с параметром  $p$ , то что бы вы выбрали в качестве оценки для  $p$ ?

2) Также оцените параметр  $p$  для 474 некурящих женщин, который также моделируется как реализация случайной выборки из геометрического распределения. Набор данных  $y_1, y_2, \dots, y_{474}$ , где  $y_j$  обозначает количество циклов для  $j$ -й некурящей женщины, представлен в таблице:

Циклы	1	2	3	4	5	6	7	8	9	10	11	12
Кол-во	198	107	55	38	18	22	7	9	5	3	6	6

3) Сравните оценки вероятности забеременеть за три или менее циклов для курящих и некурящих женщин.