

Санкт-Петербургский Политехнический университет Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчет по лабораторной работе №2 по дисциплине
«Стохастические модели и анализ данных»

Выполнил:
Чернова В.С.
Группа:
3640102/90201

Проверил:
к.ф.-м.н., доцент
Баженов А.Н.

Санкт-Петербург
2020 г

ОГЛАВЛЕНИЕ

Постановка задачи.....	3
Теоретическое введение.....	3
Ход работы	5
Заключение	7
Список литературы	7

ПОСТАНОВКА ЗАДАЧИ

Произвести дисперсионный анализ сигнала с применением критерия Фишера.

ТЕОРЕТИЧЕСКОЕ ВВЕДЕНИЕ

Критерий Фишера – это статистика, рассчитываемая в дисперсионном анализе. Этот показатель характеризует сравнение дисперсии, обусловленной вариацией самих групповых средних относительно общего среднего, с дисперсией, обусловленной вариацией признака внутри каждой отдельной группы относительно среднего по группе.

Выборочная дисперсия признака вычисляется как средняя сумма квадратов отклонений каждого значения выборки от среднего:

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{N-1} = \frac{\sum d_i^2}{N-1},$$

отсюда $\sum d_i^2 = (N-1) \cdot s^2$,

где $\sum d_i^2$ и обозначает сумму квадратов соответствующих отклонений, и, естественно, также характеризует величину вариации признака относительно среднего. Отметим, что для получения несмещенной оценки выборочной дисперсии вместо N в формуле используют величину $(N - 1)$.

Общую сумму квадратов рассчитывают по формуле:

$$\sum d_{total}^2 = (N_{total} - 1) \cdot s_{total}^2.$$

Межгрупповую сумму квадратов рассчитывают по следующей формуле:

$$\sum d_{BG}^2 = n \cdot \sum_{j=1}^N \bar{X}_j^2 - \frac{\left(\sum_{i=1}^N X_i \right)^2}{N},$$

где n – число наблюдений в каждой группе, j – знак суммирования квадратов средних по k группам, а i – суммирования значений всех наблюдений в отдельности.

Внутригрупповую сумму квадратов обычно вычисляют как разность между общей и межгрупповой суммой квадратов:

$$\sum d_{WG}^2 = \sum d_{total}^2 - \sum d_{BG}^2.$$

Далее следует определить число степеней свободы для межгрупповой и внутригрупповой вариаций наблюдений. Для некоторого множества наблюдений число степеней свободы соответствует числу не зависящих друг от друга наблюдений.

Для оценки межгрупповой вариации число степеней свободы оценивается так:

$$df_{BG} = k - 1.$$

Для оценки внутригрупповой вариации число степеней свободы оценивается так:

$$df_{WG} = N - k.$$

Межгрупповая и внутригрупповая дисперсии рассчитываются по формулам:

$$\hat{S}_{BG}^2 = \frac{\sum d_{BG}^2}{k - 1}, \quad (1)$$

$$\hat{S}_{WG}^2 = \frac{\sum d_{WG}^2}{N - k}. \quad (2)$$

Искомая статистика дисперсионного анализа – критерий Фишера – рассчитывается по формуле:

$$F = \frac{\hat{S}_{BG}^2}{\hat{S}_{WG}^2}. \quad (3)$$

ХОД РАБОТЫ

Программа была написана на языке программирования Python 3.5.

На вход подается файл с записанными сигналами, который состоит из 800 сигналов по 1024 значения. Для работы необходимо рассмотреть один из сигналов. Выберем сигнал под номером 323. Сигнал представлен на рис. 1а.

Далее необходимо сгладить выбросы. Для сглаживания выбросов был использован медианный фильтр. Сглаженный сигнал представлен на рис. 1б.

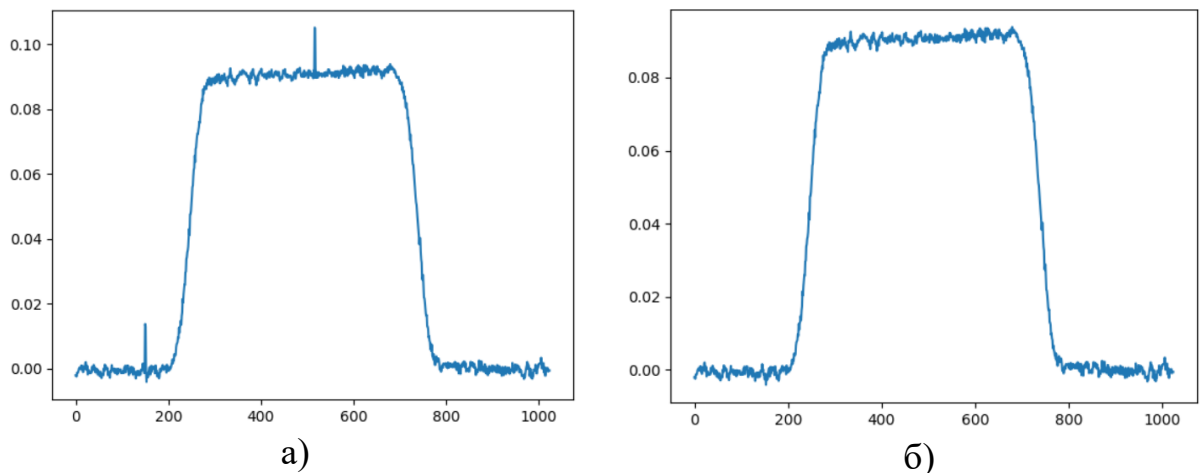


Рис. 1. Сигнал под номером 323:
а) исходный сигнал, б) сглаженный сигнал

Далее построим гистограмму сигнала, чтобы разделить сигнал на три области: сигнал, фон и переходы. Фоном является самый высокий столбец гистограммы, сигналом – столбец чуть поменьше, переходами – все остальное. Полученная гистограмма представлена на рис. 2., а сам сигнал с обозначенными на нем областями на рис. 3.

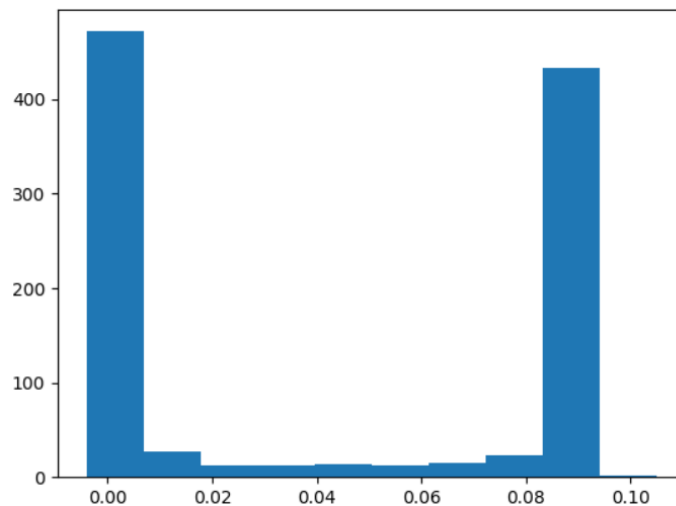


Рис. 2. Гистограмма сигнала

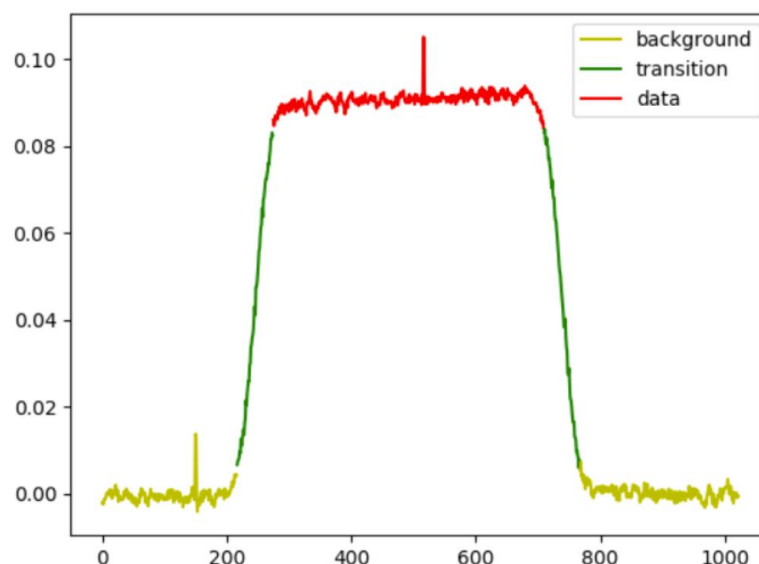


Рис. 3. Сигнал, разделенный на три области

Далее с помощью критерия Фишера определим однородность каждой части сигнала. Результаты вычисления критерия для каждого промежутка представлены в таблице 1.

Таблица 1. Расчет критерия Фишера

№	Тип	Межгрупповая дисперсия (1)	Внутригрупповая дисперсия (2)	Критерий Фишера (3)
1	Фон	2.859e-06	4.567e-06	0.626
2	Переход	0.018	3.498e-08	542207.52
3	Сигнал	3.948e-06	3.486e-06	1.132
4	Переход	0.018	7.542e-08	251147.02
5	Фон	3.744e-06	2.152e-06	1.739

Из полученных результатов можно сделать вывод, что фон и сигнал являются однородными, так как значение критерия Фишера для них близко к единице.

ЗАКЛЮЧЕНИЕ

В процессе выполнения работы был изучен и применен на практике метод дисперсионного анализа с использованием критерия Фишера. С помощью данного метода был проведен анализ сигнала.

Из полученных результатов был сделан вывод, что фон и сигнал являются однородными, так как значение критерия Фишера для них близко к единице.

Код написанной программы представлен по следующей ссылке:
https://github.com/nika2506/stochastic_labs/tree/main/Lab_2.

СПИСОК ЛИТЕРАТУРЫ

1. Гланц, С. Медико-биологическая статистика. Пер. с англ. — М.: Практика, 1998.
2. Дисперсионный анализ [Электронный ресурс], URL: https://ru.wikipedia.org/wiki/Дисперсионный_анализ. Дата обращения: 20.12.2020