Отчёт по лабораторной №2

«Применение критерия Фишера»

по дисциплине

«Стохастические модели и анализ данных»

Выполнила: Кацман Н.И.

группа: 3640102/90201

Проверил: к.ф.-м.н., доцент

Баженов А.Н.

Оглавление

[Постановка задачи 3](#_Toc67252918)

[Теоретическая часть 3](#_Toc67252919)

[Критерий фишера 3](#_Toc67252920)

[Описание критерия 3](#_Toc67252921)

[Практическая часть 4](#_Toc67252922)

[Исследуемые данные и язык реализации 4](#_Toc67252923)

[Постановка экспериментов 4](#_Toc67252924)

[Выделение фоновой и информационной частей сигнала 4](#_Toc67252925)

[Использование критерия Фишера 6](#_Toc67252926)

[Результаты 6](#_Toc67252927)

[Заключение 6](#_Toc67252928)

[Список литературы 7](#_Toc67252929)

# Постановка задачи

1. Считать набор сигналов из файла
2. Выделить в сигналах: фоновую и информационную части (и переход между ними)
3. Провести дисперсионный анализ используя критерий Фишера

# Теоретическая часть

## Критерий фишера

*Критерий Фишера* применяется для проверки равенства дисперсий двух выборок. Его относят к *критериям рассеяния*.

При проверке гипотезы положения (гипотезы о равенстве средних значений в двух выборках) с использованием [критерия Стьюдента](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9_%D0%A1%D1%82%D1%8C%D1%8E%D0%B4%D0%B5%D0%BD%D1%82%D0%B0" \o "Критерий Стьюдента) имеет смысл предварительно проверить гипотезу о равенстве дисперсий. Если она верна, то для сравнения средних можно воспользоваться более [мощным](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BE%D1%89%D0%BD%D0%BE%D1%81%D1%82%D1%8C_%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D1%8F" \o "Мощность критерия) критерием.

В [регрессионном анализе](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7" \o "Регрессионный анализ) критерий Фишера позволяет оценивать значимость линейных регрессионных моделей. В частности, он используется в [шаговой регрессии](http://www.machinelearning.ru/wiki/index.php?title=%D0%A8%D0%B0%D0%B3%D0%BE%D0%B2%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F" \o "Шаговая регрессия) для проверки целесообразности включения или исключения независимых переменных (признаков) в регрессионную модель.

В [дисперсионном анализе](http://www.machinelearning.ru/wiki/index.php?title=%D0%94%D0%B8%D1%81%D0%BF%D0%B5%D1%80%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7" \o "Дисперсионный анализ) критерий Фишера позволяет оценивать значимость факторов и их взаимодействия.

Критерий Фишера основан на дополнительных предположениях о независимости и нормальности выборок данных. Перед его применением рекомендуется выполнить [проверку нормальности](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B8_%D0%BD%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D1%81%D1%82%D0%B8).

### Описание критерия

Пусть заданы две [выборки](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0).

Обозначим через  и  [дисперсии](http://www.machinelearning.ru/wiki/index.php?title=%D0%94%D0%B8%D1%81%D0%BF%D0%B5%D1%80%D1%81%D0%B8%D1%8F_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%BE%D0%B9_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D1%8B&action=edit) выборок  и ,  и  — выборочные оценки дисперсий  и  :

;

,

где

;  — выборочные средние выборок  и .

Дополнительное предположение: выборки  и  являются [нормальными](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5). Критерий Фишера чувствителен к нарушению предположения о нормальности.

[Нулевая гипотеза](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D1%83%D0%BB%D0%B5%D0%B2%D0%B0%D1%8F_%D0%B3%D0%B8%D0%BF%D0%BE%D1%82%D0%B5%D0%B7%D0%B0)

Статистика критерия Фишера:

имеет [распределение Фишера](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%A4%D0%B8%D1%88%D0%B5%D1%80%D0%B0) с n-1 и m-1 степенями свободы. Обычно в числителе ставится большая из двух сравниваемых дисперсий. Тогда [критической областью критерия](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D1%80%D0%B8%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B0%D1%8F_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C_%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D1%8F) является правый хвост распределения Фишера, что соотвествует альтернативной гипотезе .

Критерий (при [уровне значимости](http://www.machinelearning.ru/wiki/index.php?title=%D0%A3%D1%80%D0%BE%D0%B2%D0%B5%D0%BD%D1%8C_%D0%B7%D0%BD%D0%B0%D1%87%D0%B8%D0%BC%D0%BE%D1%81%D1%82%D0%B8) α):

против альтернативы

* если или , то нулевая гипотеза  отвергается в пользу альтернативы .

против альтернативы

* если , то нулевая гипотеза  отвергается в пользу альтернативы ;

где   есть α-[квантиль](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%B2%D0%B0%D0%BD%D1%82%D0%B8%D0%BB%D1%8C" \o "Квантиль) распределения Фишера с  и   степенями свободы.

# Практическая часть

## Исследуемые данные и язык реализации

Для выполнения данный работы был выбран язык **R.**

Для исследования был использован файл с записанными сигналами. В нём 800 сигналов по 1024 значения

Работа проводилась над несколькими разными сигналами

## Постановка экспериментов

В контексте данной задачи интересно исследовать схожесть дисперсий фоновой и информационной частей сигнала. Для этого надо выделить из сигнала фон, информационную часть и переходы между ними. А после этого можно будет использовать критерий Фишера для проверки гипотезы о том, что дисперсии информационной и фоновой частей сигнала совпадают

### Выделение фоновой и информационной частей сигнала

Первым делом сигнал был сглажен, для этого был использован медианный фильтр.

После сглаживания можно предполагать, что переходы между фоном и информационной частью сигнала монотонны. Основываясь на этом предположении можно найти достаточно точные границы между фоном/переходами/информацией:

1. Постороив гистограмму можем найти наиболее редко встречающиеся значения – они будут соответстовать переходам. Возьмём по одному значению с каждого из переходв (их два: фон -> информация и информация -> фон)

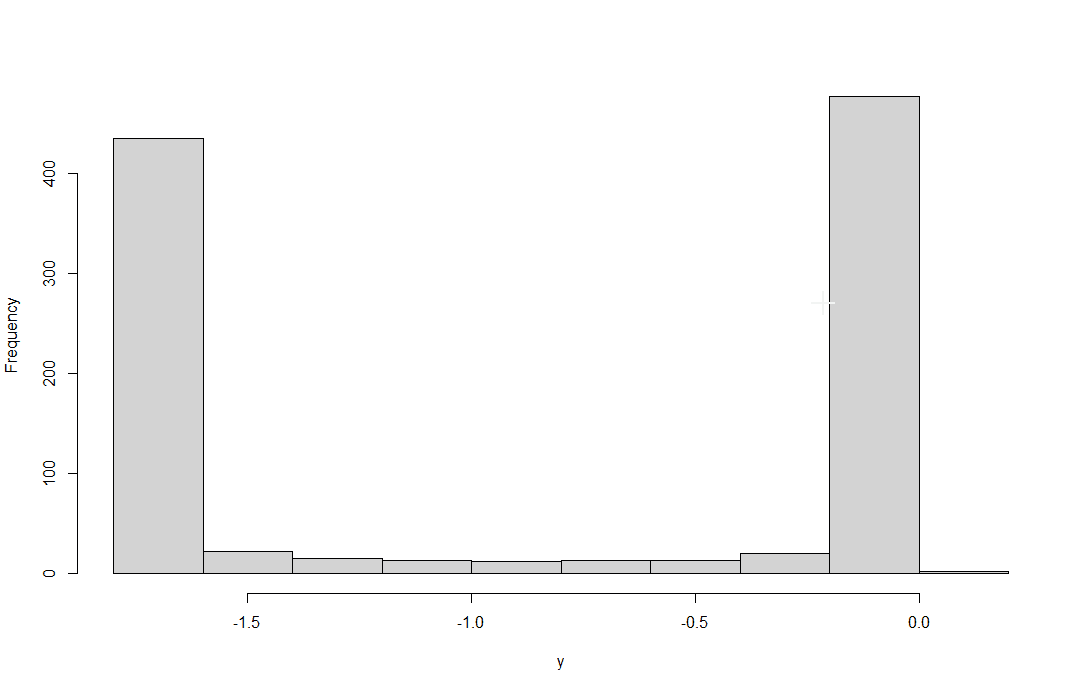


Рисунок 1 Гистограмма значений сигнала 33

1. Для каждого из выбранных в предыдущем пункте значений будем двигаться влево и вправо проверяя не изменится ли монотонность. Точка смены монотонности – граница между фоном/информацией и переходом (это может быть не настоящей точкой перехода между фоном/информацией и переходом, но точкой, находящейся достаточно близко к ней)

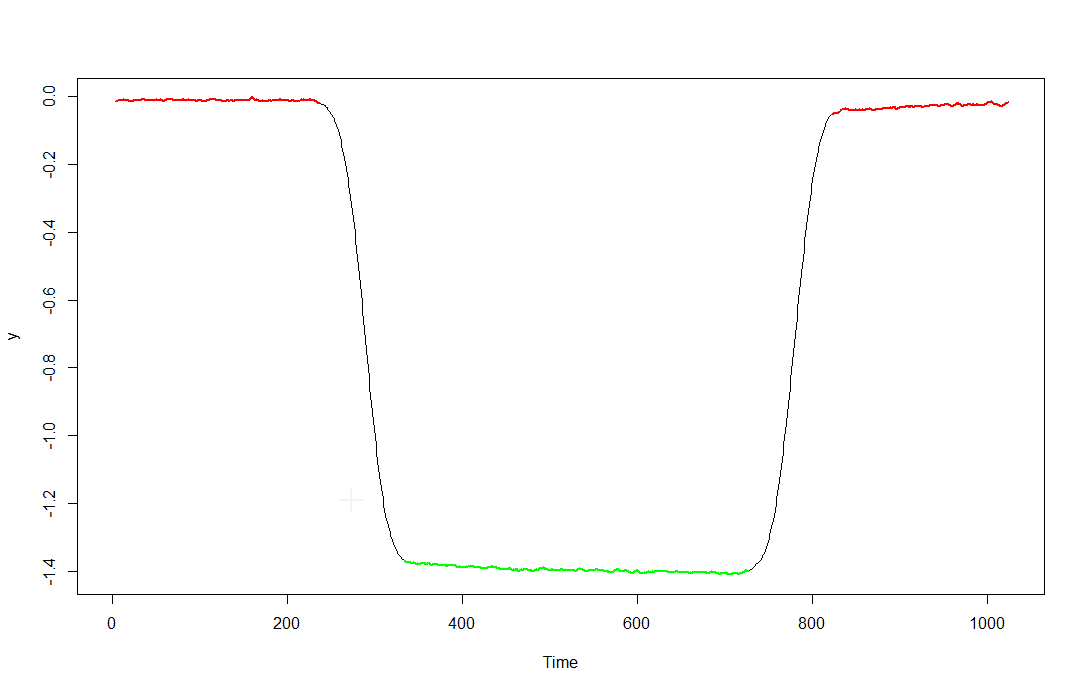


Рисунок 2 Сигнал 33. Красный - фон, Зелёный - Информация

В результате для каждого из сигналов были определены 5 интервалов:

* левая часть фона
* переход от фона к информационной части сигнала
* информационная часть сигнала
* переход от информационной части сигнала к фоновой
* правая часть фона

### Использование критерия Фишера

Интервалы перехода от фоновой к информационной части сигнала и обратно очевидно далеки от нормального распределения, так что они в данной части не рассматриваются

Интересно попробовать выяснить является ли дисперсия в информационной части сигналов такой же как и в фоновой. Для этого «склеим» правую и левую части фона и посчитаем критерий Фишера для двух выборок: склееный фон и информационная часть сигнала

## Результаты

В таблице ниже приведены посчитанные критерии Фишера для оценки вероятности того, что гипотеза о равенстве дисперсий верна (уровень значимости был взят равням 0,05)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер сигнала | 78 | 118 | 198 | 219 | 512 |
| Критерий Фишера | 1.0486 | 0.73058 | 0.76281 | 0.95389 | 1.0639 |
| 95% доверительный интервал | [0.810809, 1.2024887] | [0.811559, 1.201395] | [0.8104348, 1.2147390] | [0.8229711, 1.200764] | [0.8182286,  1.2029646] |

# Заключение

В ходе работы был написан код, позволяющий для заданного номера сигнала из входного файла:

* Получить данные этого сигнала
* Выделить в сигнале фоновую и иноформационные части
* Проверить гипотезу о том что дисперсии фоновой и информационной частей сигнала совпадают

Из данных, приведённых в таблице выше (и вычисленных, но не вошедших в эту таблицу), можно сделать вывод, что скорее всего дисперсии фоновой и информационной части сигналов из входного файла совпадают.

Файл кода и этот отчёт можно посмотреть [здесь](https://github.com/Hadegda/Stochastic-labs/blob/main/lab2_Fisher/stochastic2.R).

# Список литературы

1. Критерий Фишера [Электронный ресурс]   
   url: http://www.machinelearning.ru/wiki/index.php?title=Критерий\_Фишера  
   Дата обращения: 16.03.21