РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

Факультет физико-математических и естественных наук Кафедра прикладной информатики и теории вероятностей

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №7

<u>дисциплина: «Компьютерный практикум по статистическому</u> анализу данных»

Работу выполнил:

Снимщиков Иван Игоревич

Группа: НПИбд-02-21

MOCKBA

2024 г.

Цели работы: Основной целью работы является специализированных пакетов Julia для обработки данных.

Ход работы:

Первым делом я установил нужные пакеты для выполнения лабораторной работы:

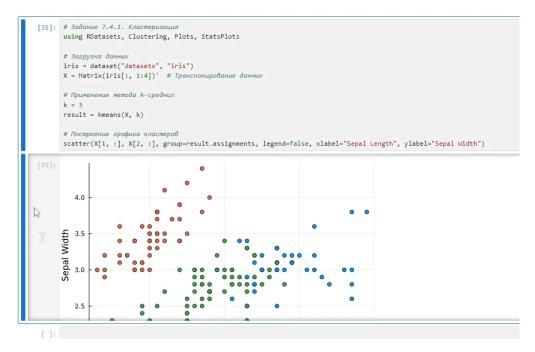
- Clustering
- CSV
- DataFrames
- RDatasets
- FileIO
- NearestNeighbors
- MultivariateStats
- PyCall
- Conda
- BenchmarkTools

```
(@v1.10) pkg> add Clustering
    Updating registry at `F:\Users\
    Resolving package versions...
```

```
(@v1.10) pkg> add CSV
Resolving package versions...
Installed PooledArrays — v1.4.3
Installed FilePathsBase — v0.9.22
Installed SentinelArrays — v1.4.8
Installed WeakRefStrings — v1.4.2
Installed InlineStrings — v1.4.2
Installed WorkerUtilities — v1.6.1
```

```
recompiled. 3 skipped during auto due to pre
@v1.10) pkg> add DataFrames
  Resolving package versions...
@v1.10) pkg> add RDatasets
 Resolving package versions...
Installed CategoricalArrays - v0.10.8
 Installed TZJData V1.3.1+2024b
Installed Mocking V0.8.1
Installed RData V0.8.3
Installed TimeZones V1.20.0
  Installed FileIO — v1.16.5
 Progress [======== ] 6/7
@v1.10) pkg> add FileIO
  Resolving package versions...
@v1.10) pkg> add NearestNeighbors
  Resolving package versions...
(@v1.10) pkg> add MultivariateStats
  Resolving package versions...
  @v1.10) pkg> add PyCall
  Resolving package versions...
(@v1.10) pkg> add Conda
  Resolving package versions...
(@v1.10) pkg> add BenchmarkTools
  Resolving package versions...
  Installed BenchmarkTools - v1.5.0
```

Задание 7.4.1 Кластеризация



Задание 7.4.2 Регрессия

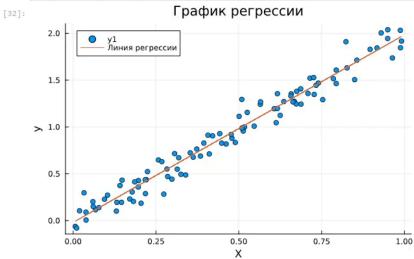
```
# MHK-оценка
X_aug = hcat(ones(1000), X)
beta = X_aug \ y

# Сравнение с llsq us MultivariateStats
model = fit(LinearModel, X, y)
coeff_llsq = coef(model)

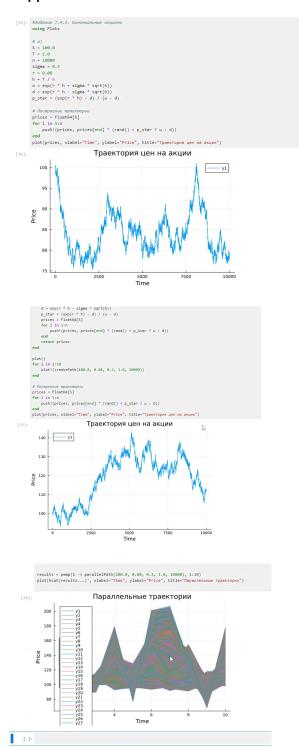
# Сравнение с GLM
glm_model = lm(@formula(y ~ x1 + x2 + x3), DataFrame(y=y, x1=X[:,1], x2=X[:,2], x3=X[:,3]))
coeff_glm = coef(glm_model)

# Часть 2
X = rand(100)
y = 2 .* X + 0.1 .* randn(100)

scatter(X, y, xlabel="X", ylabel="y", title="График регрессии")
coeff = [ones(length(X)) X] \ y
plot!(X, coeff[2] .* X .+ coeff[1], label="Линия регрессии")
```



Задание 7.4.3



Вывод: Я освоил специализированные пакетов Julia для обработки данных.