

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

Задание 1.
Поисковый анализ, описательная статистика

Отчёт
О выполненном задании

Выполнил:
студент 523 группы
Латыпов Ш. И.
latypovshamil2001@gmail.com

Москва
2023

Описание задачи

Выполняется на данных "Термодинамика". В этом задании нет разделения признаков на предикторы и целевые.

Для каждого признака отдельно указать шкалу. Для каждого признака отдельно привести описательные статистики.

- набор описательных статистик зависит от шкалы
- нарисовать распределение

Найти функциональные (точные, а не статистические) связи между признаками.

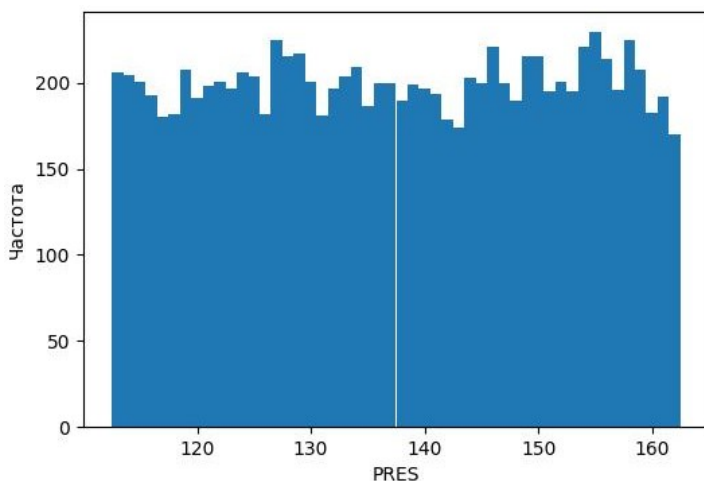
- речь о довольно простых связях, раскрывающих смысл признаков

Данные из файла "ТемпИзменяется.data"

Характеристики данных

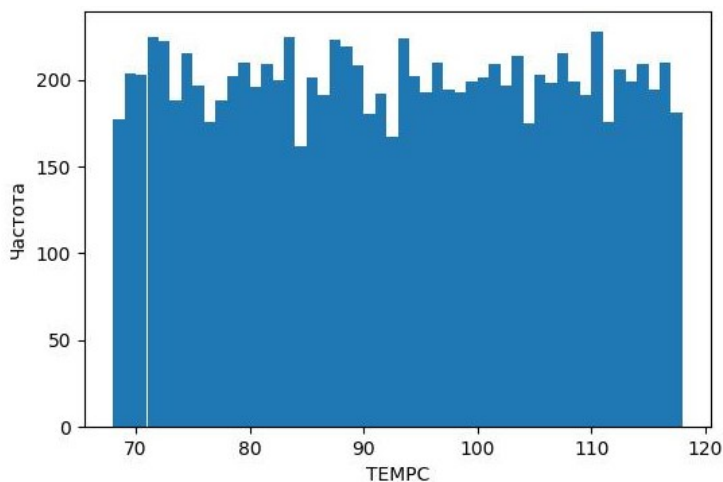
В данной секции приведены графики и описательные статистики для столбцов из данных файла "ТемпИзменяется.data".

1. PRES - Давление



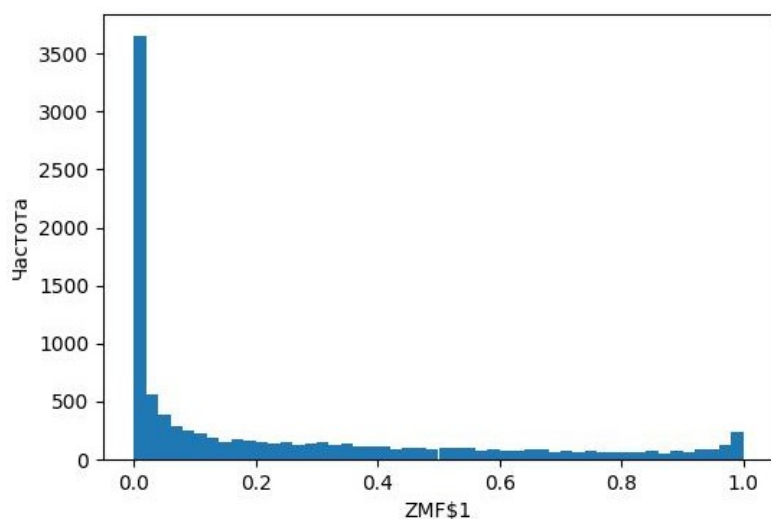
```
PRES
count    10000.000000
mean      137.583220
std       14.412738
min       112.501200
25%       125.138475
50%       137.536800
75%       150.186400
max       162.494900
Name: PRES, dtype: float64
Медиана: 137.5368
Дисперсия: 207.7270033612751
Коэффициент вариации: 0.10475650681129405
```

2. TEMPC - Температура



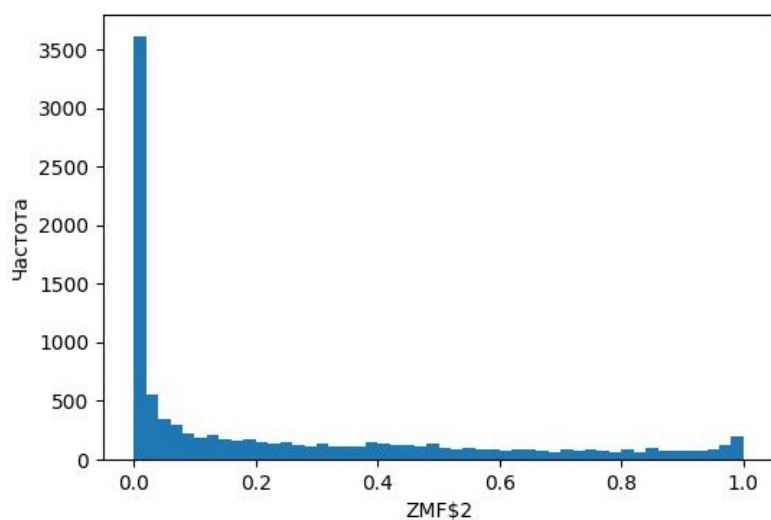
```
TEMPC
count    10000.000000
mean      92.971691
std       14.431558
min       68.013000
25%       80.498000
50%       93.065000
75%       105.510250
max       117.991000
Name: TEMPC, dtype: float64
Медиана: 93.065
Дисперсия: 208.269874792407
Коэффициент вариации: 0.1552252955833905
```

3. ZMF\$1 - Общая молярная концентрация компонента 1



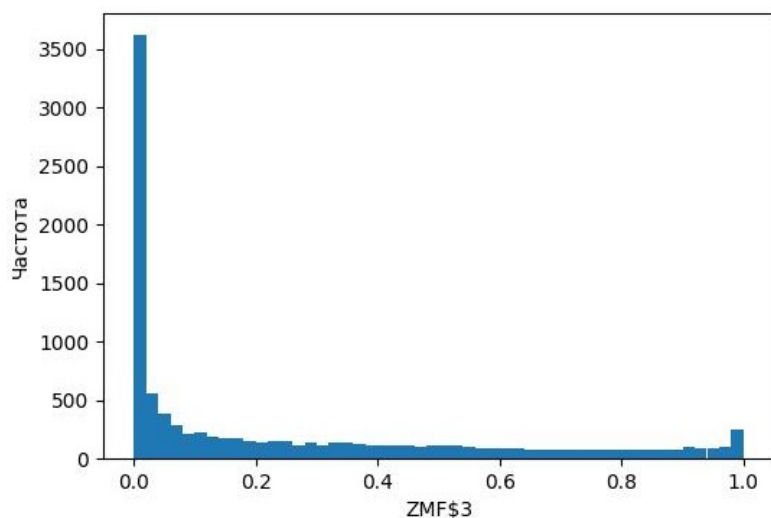
```
ZMF$1
count      10000.000000
mean        0.246731
std         0.306152
min         0.000000
25%         0.002847
50%         0.086733
75%         0.424530
max         1.000000
Name: ZMF$1, dtype: float64
Медиана: 0.08673296999999999
Дисперсия: 0.09372908434335839
Коэффициент вариации: 1.2408332746450976
```

4. ZMF\$2 - Общая молярная концентрация компонента 2



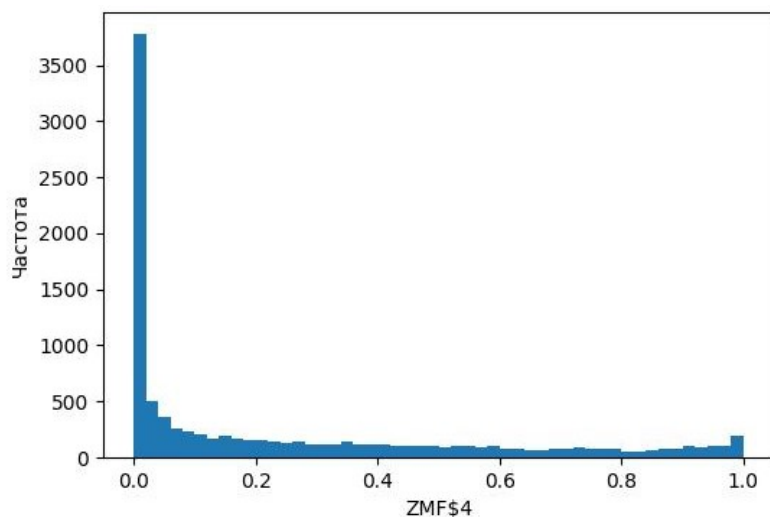
```
ZMF$2
count      10000.000000
mean        0.253145
std         0.306498
min         0.000000
25%         0.003224
50%         0.097082
75%         0.445326
max         0.999986
Name: ZMF$2, dtype: float64
Медиана: 0.097082166
Дисперсия: 0.09394084751785838
Коэффициент вариации: 1.210760385627354
```

5. ZMF\$3 - Общая молярная концентрация компонента 3



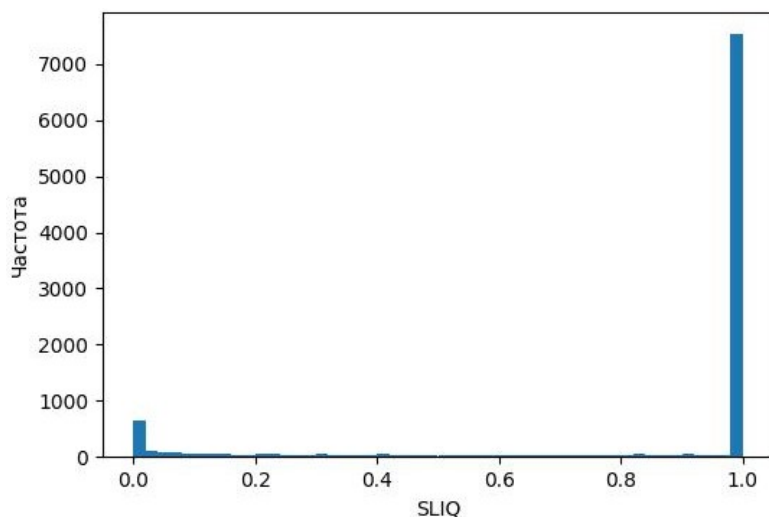
```
ZMF$3
count      10000.000000
mean        0.253402
std         0.309071
min         0.000000
25%         0.003417
50%         0.094978
75%         0.447298
max         0.999999
Name: ZMF$3, dtype: float64
Медиана: 0.0949779905
Дисперсия: 0.09552470510851085
Коэффициент вариации: 1.219684414396855
```

6. ZMF\$4 - Общая молярная концентрация компонента 4



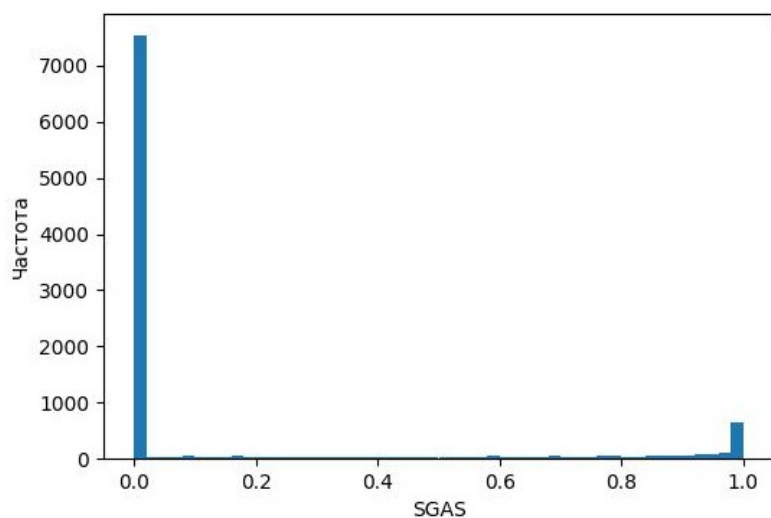
```
ZMF$4
count      10000.000000
mean        0.246722
std         0.305948
min         0.000000
25%         0.002701
50%         0.087613
75%         0.431056
max         0.999924
Name: ZMF$4, dtype: float64
Медиана: 0.087612899
Дисперсия: 0.09360390437230814
Коэффициент вариации: 1.240049905674024
```

7. SLIQ - Насыщение жидкости



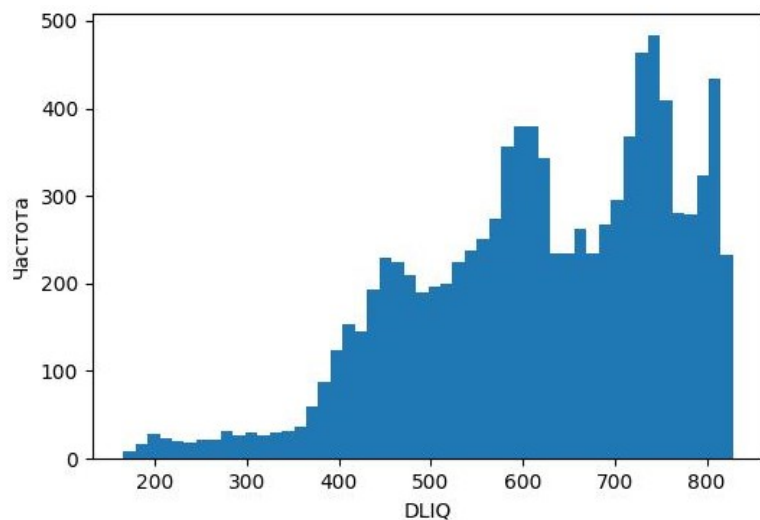
```
SLIQ
count      10000.000000
mean        0.835697
std         0.329341
min         0.000000
25%         1.000000
50%         1.000000
75%         1.000000
max         1.000000
Name: SLIQ, dtype: float64
Медиана: 1.0
Дисперсия: 0.1084654800277154
Коэффициент вариации: 0.3940911790215165
```

8. SGAS - Насыщение газа



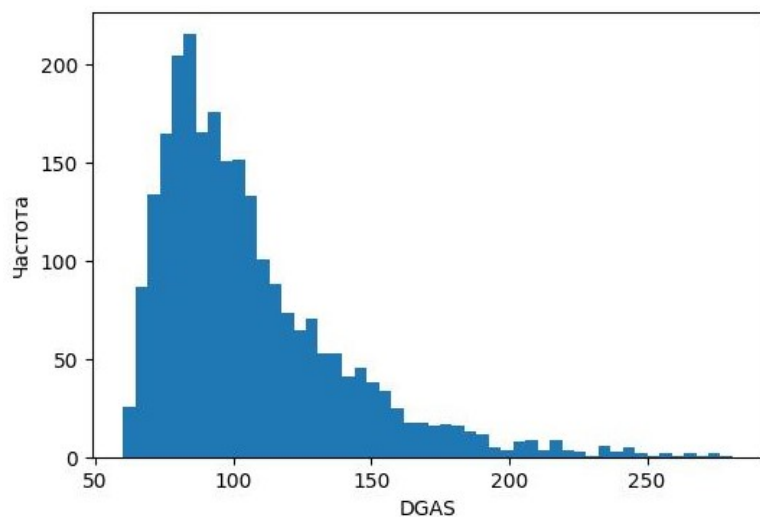
```
SGAS
count      10000.000000
mean        0.164303
std         0.329341
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max         1.000000
Name: SGAS, dtype: float64
Медиана: 0.0
Дисперсия: 0.10846548002771543
Коэффициент вариации: 2.0044783357913576
```

9. DLIQ - Плотность жидкости



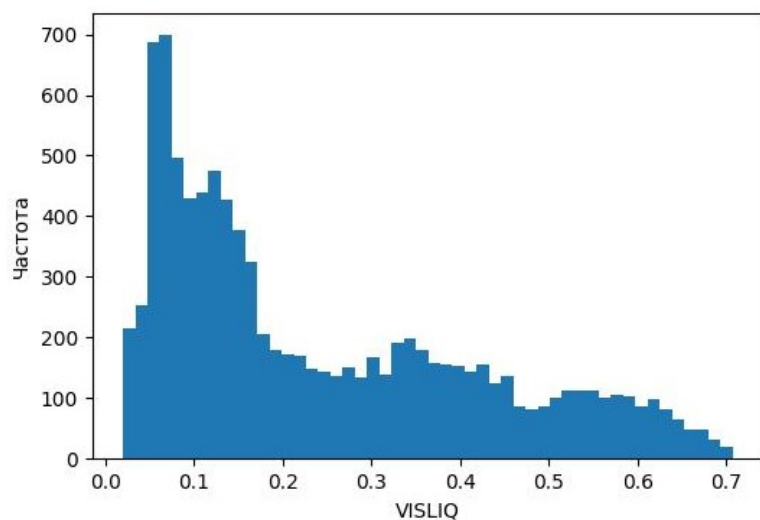
DLIQ
count 9633.000000
mean 620.198601
std 138.360571
min 165.668900
25% 524.981600
50% 629.035800
75% 736.523400
max 828.607700
Name: DLIQ, dtype: float64
Медиана: 629.0358
Дисперсия: 19143.64763170985
Коэффициент вариации: 0.22309075023641178

10. DGAS - Плотность газа



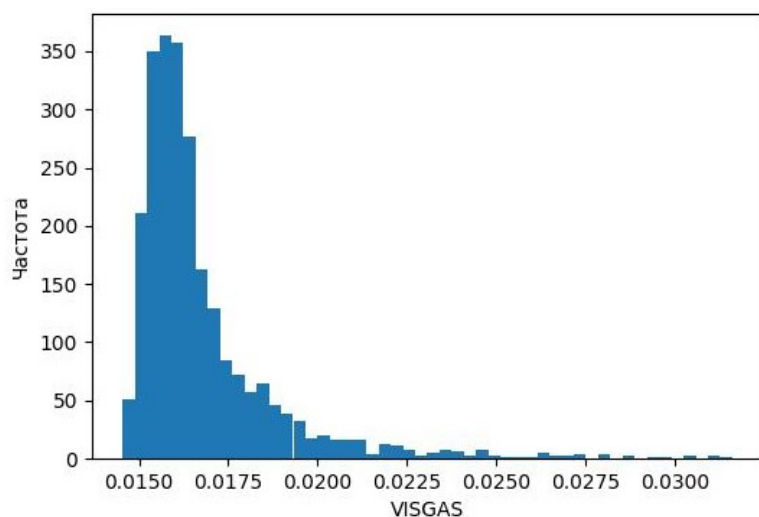
DGAS
count 2483.000000
mean 106.540499
std 34.544143
min 60.023900
25% 82.205950
50% 97.091700
75% 121.278500
max 281.007500
Name: DGAS, dtype: float64
Медиана: 97.0917
Дисперсия: 1193.2977996526124
Коэффициент вариации: 0.3242348520258719

11. VISLIQ - Вязкость жидкости



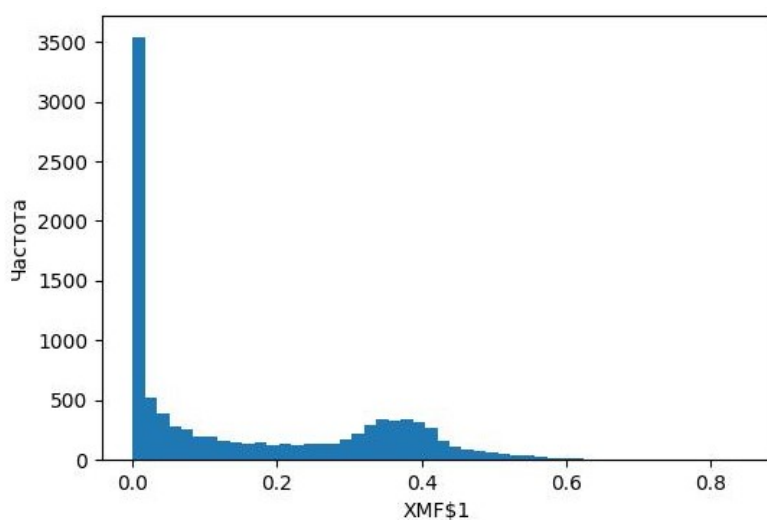
VISLIQ
count 9633.000000
mean 0.241949
std 0.179903
min 0.020040
25% 0.090610
50% 0.170800
75% 0.372200
max 0.707700
Name: VISLIQ, dtype: float64
Медиана: 0.1708
Дисперсия: 0.032364981555879406
Коэффициент вариации: 0.7435576170640873

12. VISGAS - Вязкость жидкости



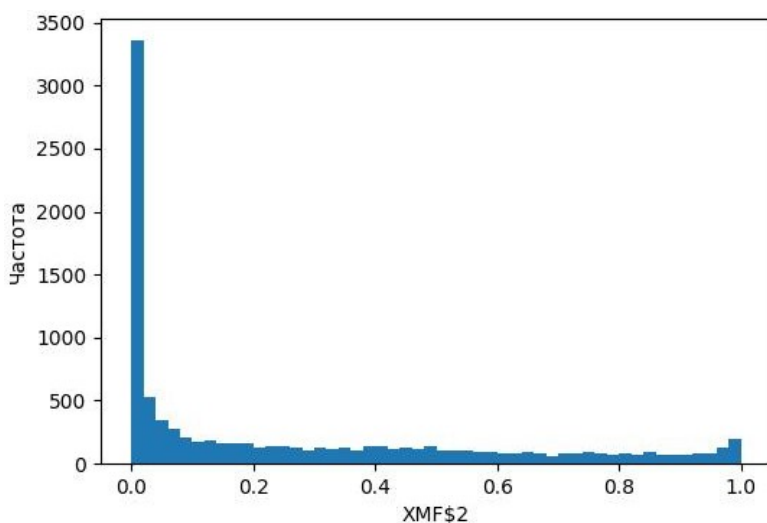
VISGAS
count 2483.000000
mean 0.016788
std 0.002093
min 0.014540
25% 0.015570
50% 0.016140
75% 0.017175
max 0.031610
Name: VISGAS, dtype: float64
Медиана: 0.01614
Дисперсия: 4.382116162637604e-06
Коэффициент вариации: 0.12469437964032014

13. XMF\$1 - Молярная концентрация компонента 1 в жидком состоянии



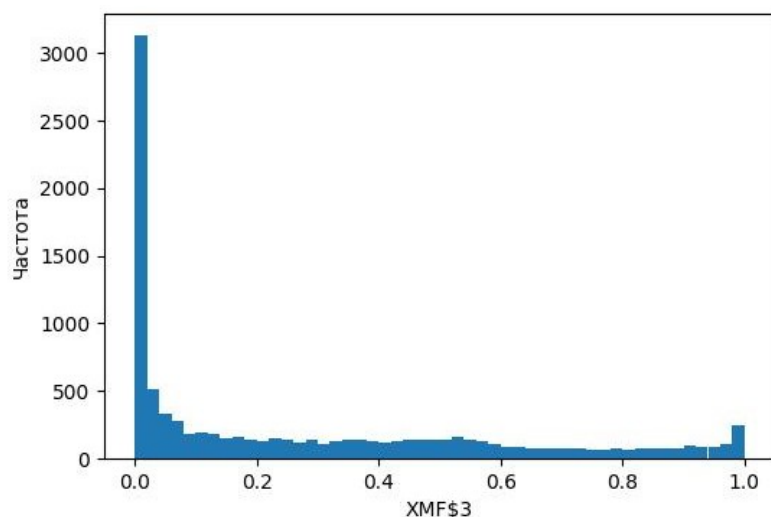
XMF\$1
count 9633.000000
mean 0.157884
std 0.170674
min 0.000000
25% 0.002373
50% 0.074389
75% 0.329771
max 0.842648
Name: XMF\$1, dtype: float64
Медиана: 0.074388686
Дисперсия: 0.029129688625003744
Коэффициент вариации: 1.0810073186202844

14. XMF\$2 - Молярная концентрация компонента 2 в жидком состоянии



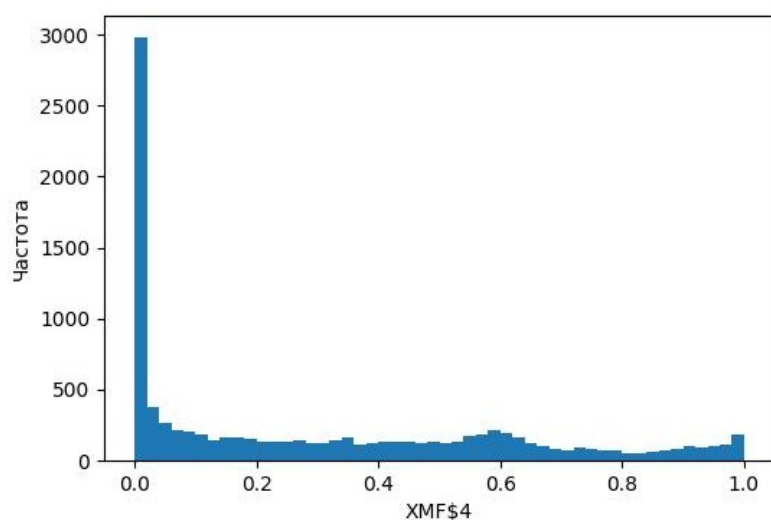
XMF\$2
count 9633.000000
mean 0.262969
std 0.309608
min 0.000000
25% 0.003811
50% 0.111961
75% 0.469025
max 0.999986
Name: XMF\$2, dtype: float64
Медиана: 0.111960956
Дисперсия: 0.09585734448416063
Коэффициент вариации: 1.17735580724908

15. XMF\$3 - Молярная концентрация компонента 3 в жидком состоянии



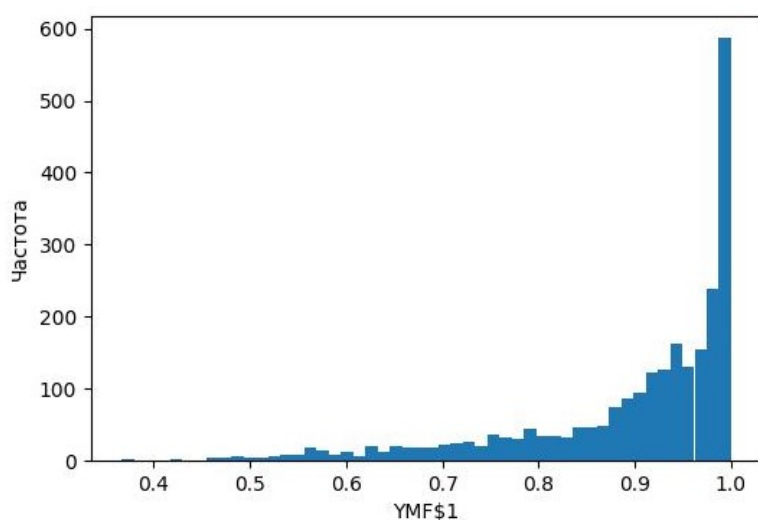
```
XMF$3
count      9633.000000
mean        0.278136
std         0.311333
min         0.000000
25%         0.005897
50%         0.140659
75%         0.497820
max         0.999999
Name: XMF$3, dtype: float64
Медиана: 0.140659134
Дисперсия: 0.09692819018277747
Коэффициент вариации: 1.119356756793551
```

16. XMF\$4 - Молярная концентрация компонента 4 в жидком состоянии



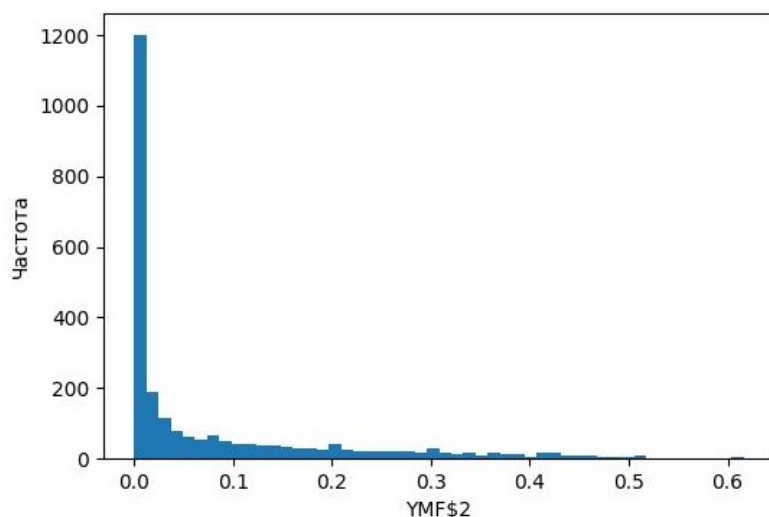
```
XMF$4
count      9633.000000
mean        0.301011
std         0.311298
min         0.000000
25%         0.006924
50%         0.194563
75%         0.555335
max         0.999924
Name: XMF$4, dtype: float64
Медиана: 0.194563119
Дисперсия: 0.09690668785255921
Коэффициент вариации: 1.0341769493771866
```

17. YMF\$1 - Молярная концентрация компонента 1 в газообразном состоянии



```
YMF$1
count      2483.000000
mean        0.894771
std         0.121404
min         0.366777
25%         0.850879
50%         0.938783
75%         0.985841
max         1.000000
Name: YMF$1, dtype: float64
Медиана: 0.938783185
Дисперсия: 0.014738862323736899
Коэффициент вариации: 0.13568132683505352
```

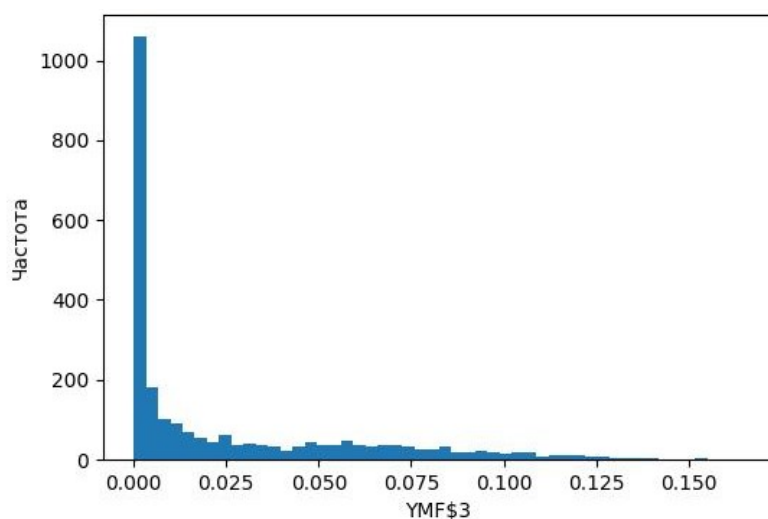
18. YMF\$2 - Молярная концентрация компонента 2 в газообразном состоянии



```

YMF$2
count      2483.000000
mean        0.078875
std         0.121086
min         0.000000
25%         0.000424
50%         0.014588
75%         0.111353
max         0.616548
Name: YMF$2, dtype: float64
Медиана: 0.014588491
Дисперсия: 0.014661890898673058
Коэффициент вариации: 1.5351604935433043
    
```

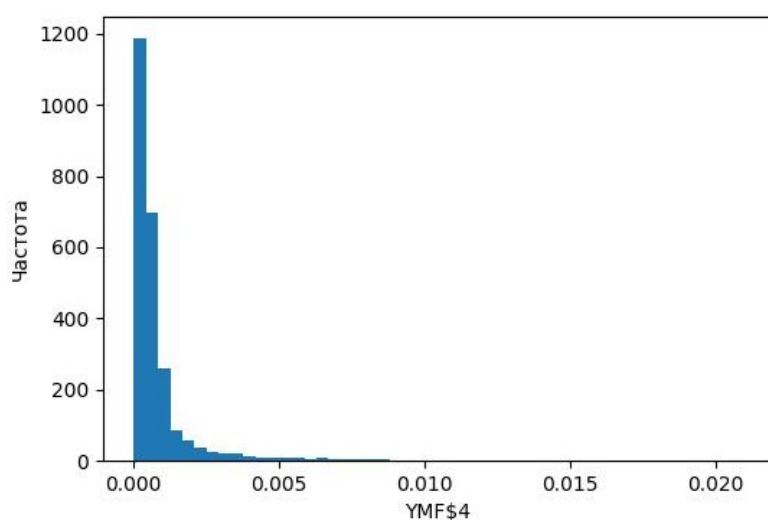
19. YMF\$3 - Молярная концентрация компонента 3 в газообразном состоянии



```

YMF$3
count      2483.000000
mean        0.025527
std         0.034145
min         0.000000
25%         0.000267
50%         0.006574
75%         0.045858
max         0.164870
Name: YMF$3, dtype: float64
Медиана: 0.00657443
Дисперсия: 0.0011658637543114753
Коэффициент вариации: 1.3375992499774778
    
```

20. YMF\$4 - Молярная концентрация компонента 4 в газообразном состоянии



```

YMF$4
count      2483.000000
mean        0.000827
std         0.001589
min         0.000000
25%         0.000153
50%         0.000447
75%         0.000818
max         0.020949
Name: YMF$4, dtype: float64
Медиана: 0.000446577
Дисперсия: 2.5243429562958873e-06
Коэффициент вариации: 1.921694029181077
    
```


Корреляционная зависимость

Для всех данных исследована корреляционная зависимость, используя стандартные функции библиотеки pandas.

```
import pandas as pd
df = pd.read_csv('ТемпИзменяется.data', sep='\s+')

# Вычисляем матрицу корреляции для всех пар числовых переменных
correlation_matrix = df.corr()

import seaborn as sns
import matplotlib.pyplot as plt

# Построение тепловой карты матрицы корреляции
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm')
plt.savefig("corr_table.jpg")
```

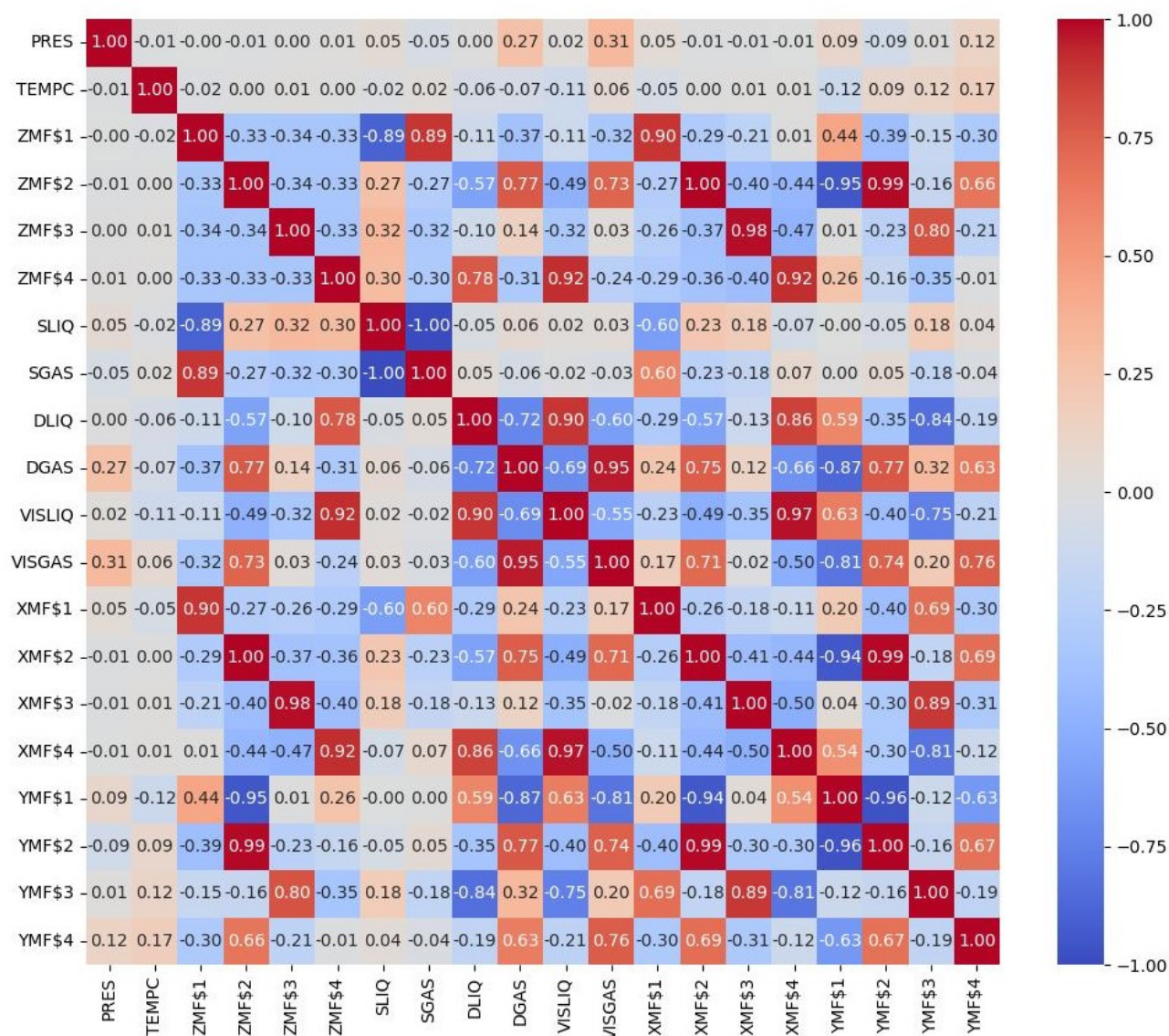


Рис. 1: Таблица корреляционной зависимости

Функциональная зависимость

После визуального осмотра данных, было замечено, что сумма параметров ZMF, XMF и YMF по всем компонентам равна 1, т.е.:

$$XMF\$1 + XMF\$2 + XMF\$3 + XMF\$4 = 1,$$

$$YMF\$1 + YMF\$2 + YMF\$3 + YMF\$4 = 1,$$

$$ZMF\$1 + ZMF\$2 + ZMF\$3 + ZMF\$4 = 1.$$

Для поиска функциональной зависимости был проведен регрессионный анализ. С помощью линейной регрессии можно попытаться подобрать модели, которые описывают зависимости между переменными.

Для сравнения были выбраны некоторые пары переменных:

- **Давление (PRES) и Температура (TEMPC):**

Можно исследовать, как температура влияет на давление смеси. Если данные соответствуют закону Гей-Люссака для идеальных газов, где давление пропорционально температуре при постоянном объеме, можно ожидать положительную корреляцию.

- **Плотность (DLIQ или DGAS) и Давление (PRES):**

Плотность может быть функцией давления, особенно в случае сжимаемых сред, таких как газы.

- **Вязкость (VISLIQ или VISGAS) и Температура (TEMPC):**

Вязкость жидкостей и газов часто зависит от температуры, при этом вязкость жидкостей обычно уменьшается с повышением температуры.

- **Молярная концентрация компонентов (XMF_c, YMF_c) и Давление (PRES):**

Концентрации фаз могут изменяться в зависимости от давления, что может быть важно при изучении фазовых переходов.

- **Насыщенность фаз (SLIQ или SGAS) и Давление (PRES):**

Изменение давления может влиять на равновесие между жидкой и газовой фазами, что в свою очередь влияет на насыщенность.

Линейная зависимость искалась через библиотеку statsmodels.

```
import statsmodels.api as sm

# Предположим, что у вас есть DataFrame df с колонками 'PRES' и 'TEMPC'
# 'PRES' будет зависимой переменной, а 'TEMPC' - независимой

# Добавляем константу к независимым переменным для интерцепта
X = sm.add_constant(df['TEMPC']) # TEMPC - независимая переменная
y = df['PRES']                  # PRES - зависимая переменная

# Построение модели линейной регрессии
model = sm.OLS(y, X).fit()

# Просмотр статистики модели
print(model.summary())
```

По результатам этого анализа, было выявлено, что:

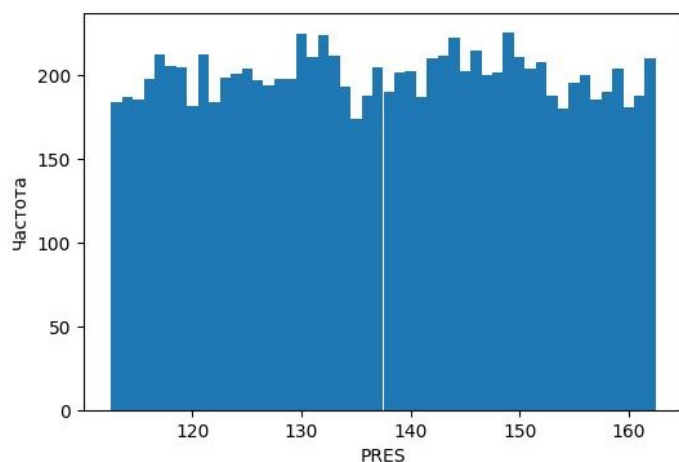
- Между PRES и TEMPC нет линейной зависимости.
- Между DLIQ и PRES нет линейной зависимости.
- Между DGAS и PRES есть очень незначительная линейная зависимость. R-squared принимает значение 0.079.
- Между VISLIQ и TEMPC нет линейной зависимости.
- Между VISGAS и TEMPC нет линейной зависимости.
- XMF и PRES: Только первый компонент имеет линейную зависимость от давления, со значением R-squared 0.311.
- YMF и PRES: Ни у одного из компонентов нет линейной зависимости от PRES, для всех компонентов значение R-squared не превышает 0.01.
- SLIQ и SGAS - линейной зависимости с давлением PRES нет.

Данные из файла "ТемпПостоянная.data"

Характеристики данных

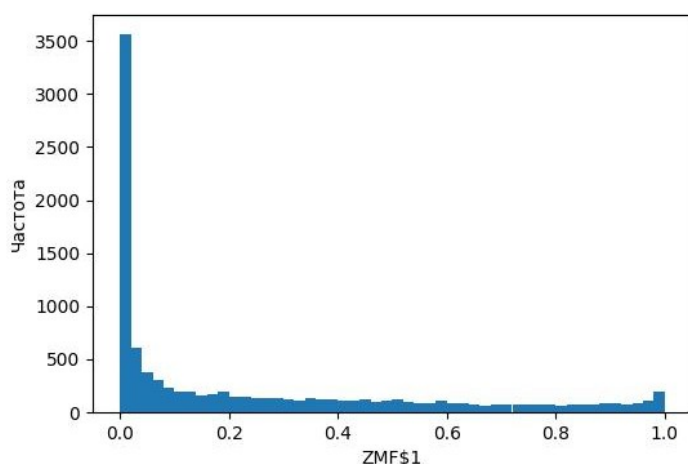
В данной секции приведены графики и описательные статистики для столбцов из данных файла "ТемпПостоянная.data".

1. PRES - Давление



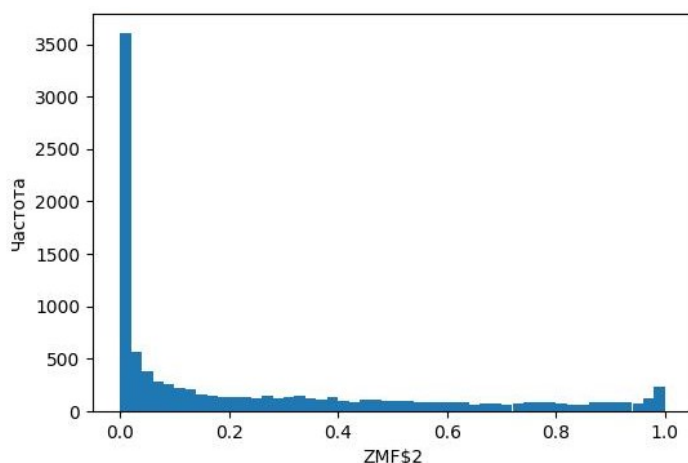
PRES
count 10000.000000
mean 137.529504
std 14.306405
min 112.509100
25% 125.200150
50% 137.598750
75% 149.699350
max 162.498000
Name: PRES, dtype: float64
Медиана: 137.59875
Дисперсия: 204.67321131599786
Коэффициент вариации: 0.10402425756221827

2. ZMF\$1 - Общая молярная концентрация компонента 1



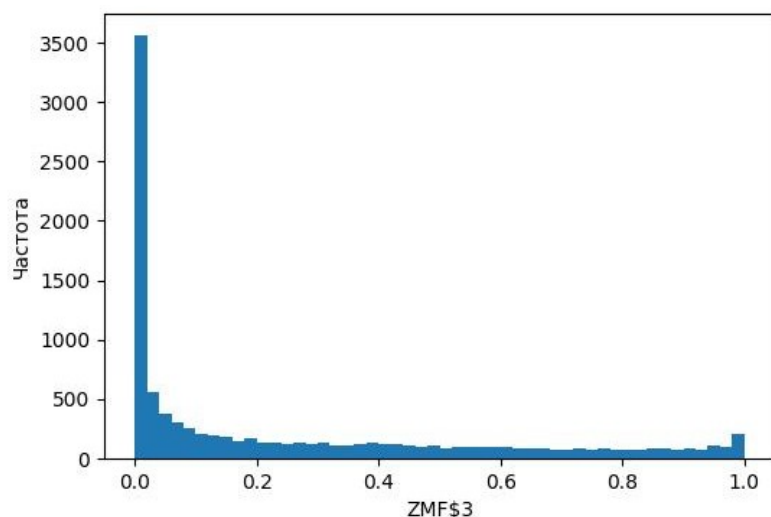
ZMF\$1
count 10000.000000
mean 0.249655
std 0.304854
min 0.000000
25% 0.003188
50% 0.093278
75% 0.439522
max 1.000000
Name: ZMF\$1, dtype: float64
Медиана: 0.093277782
Дисперсия: 0.09293615312873527
Коэффициент вариации: 1.2211002604499697

3. ZMF\$2 - Общая молярная концентрация компонента 2



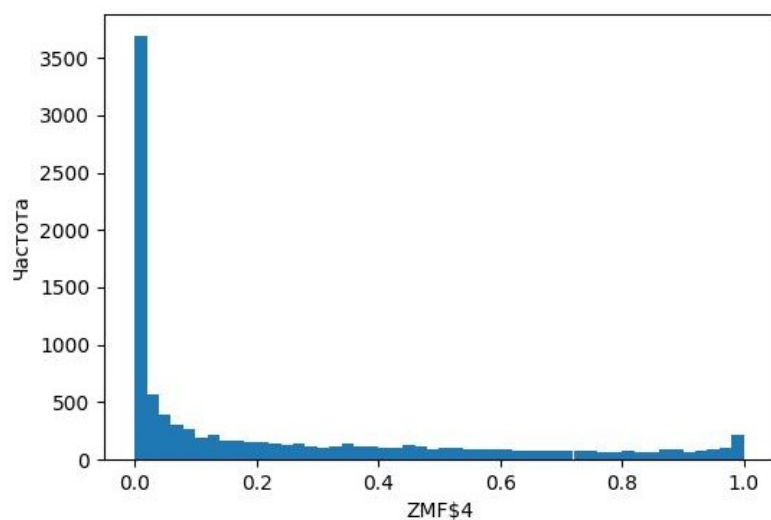
ZMF\$2
count 10000.000000
mean 0.251731
std 0.309137
min 0.000000
25% 0.003112
50% 0.092001
75% 0.441189
max 0.999995
Name: ZMF\$2, dtype: float64
Медиана: 0.09200103949999999
Дисперсия: 0.09556564434651597
Коэффициент вариации: 1.2280463449638384

4. ZMF\$3 - Общая молярная концентрация компонента 3



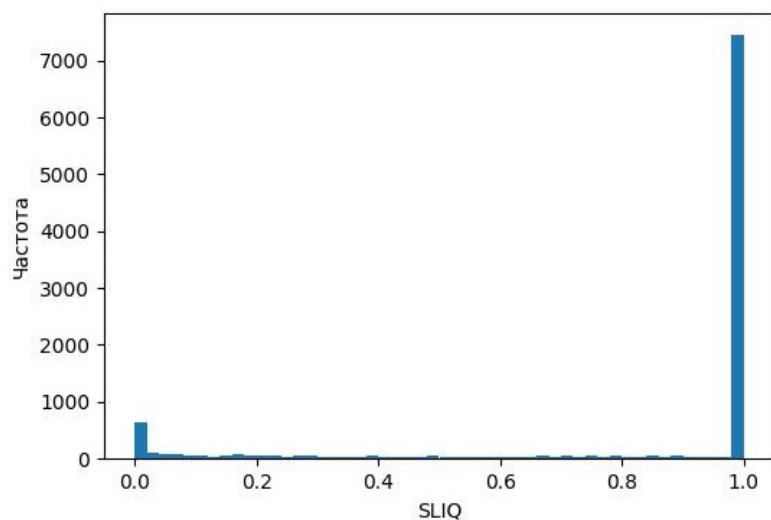
ZMF\$3
count 10000.000000
mean 0.253819
std 0.307950
min 0.000000
25% 0.003316
50% 0.095318
75% 0.446651
max 0.999992
Name: ZMF\$3, dtype: float64
Медиана: 0.095318283
Дисперсия: 0.09483291427769842
Коэффициент вариации: 1.2132621773511694

5. ZMF\$4 - Общая молярная концентрация компонента 4



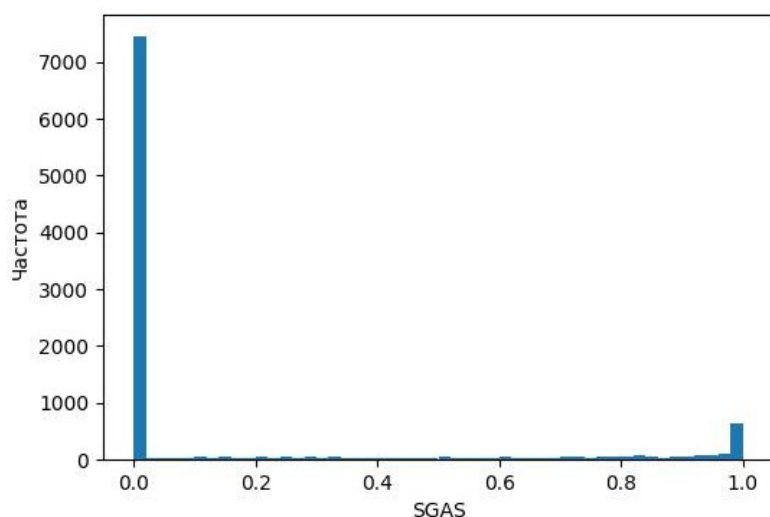
ZMF\$4
count 10000.000000
mean 0.244794
std 0.305536
min 0.000000
25% 0.002815
50% 0.083368
75% 0.429631
max 0.999991
Name: ZMF\$4, dtype: float64
Медиана: 0.0833678665
Дисперсия: 0.09335201536404249
Коэффициент вариации: 1.2481313335750264

6. SLIQ - Насыщение жидкости



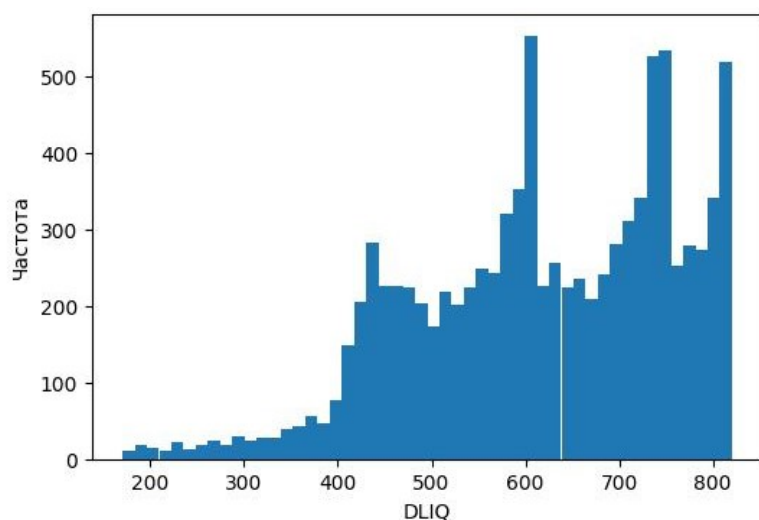
SLIQ
count 10000.000000
mean 0.832826
std 0.328466
min 0.000000
25% 0.952835
50% 1.000000
75% 1.000000
max 1.000000
Name: SLIQ, dtype: float64
Медиана: 1.0
Дисперсия: 0.10789017478309514
Коэффициент вариации: 0.39439997011155226

7. SGAS - Насыщение газа



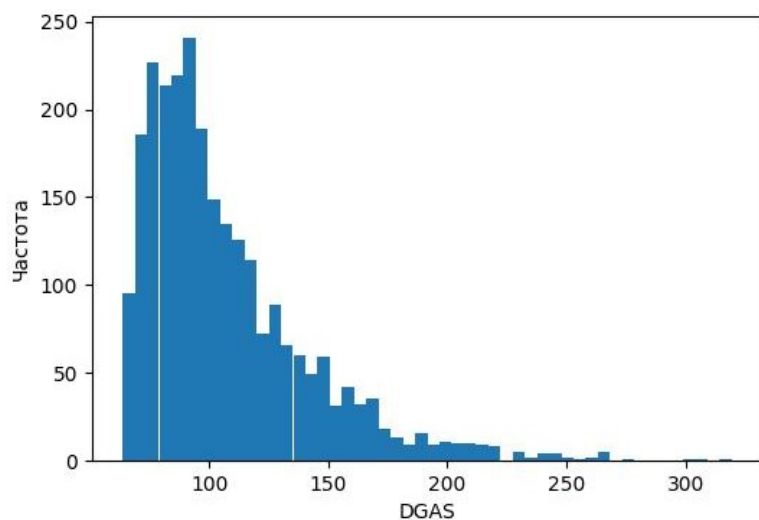
```
SGAS
count      10000.000000
mean         0.167174
std          0.328466
min           0.000000
25%           0.000000
50%           0.000000
75%          0.047165
max           1.000000
Name: SGAS, dtype: float64
Медиана: 0.0
Дисперсия: 0.10789017478309514
Коэффициент вариации: 1.9648129777953396
```

8. DLIQ - Плотность жидкости



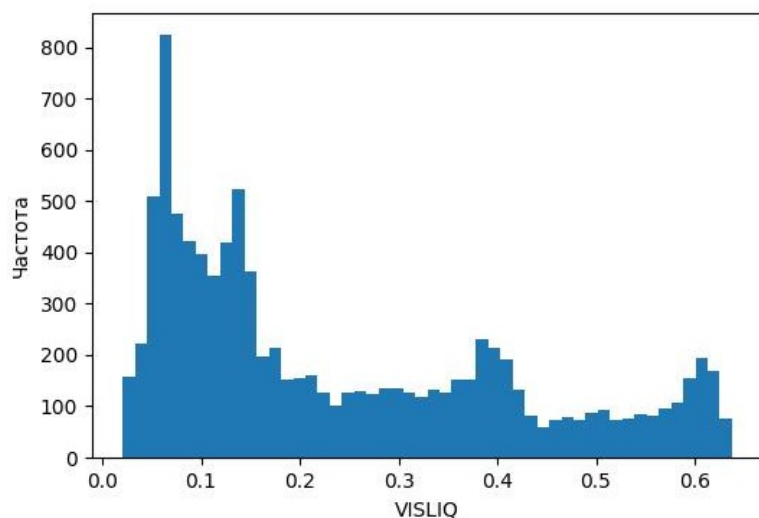
```
DLIQ
count      9652.000000
mean        619.146622
std         136.784476
min         170.470600
25%         519.379775
50%         625.763250
75%         737.950125
max         820.266200
Name: DLIQ, dtype: float64
Медиана: 625.76325
Дисперсия: 18709.992947635492
Коэффициент вариации: 0.2209242066172669
```

9. DGAS - Плотность газа



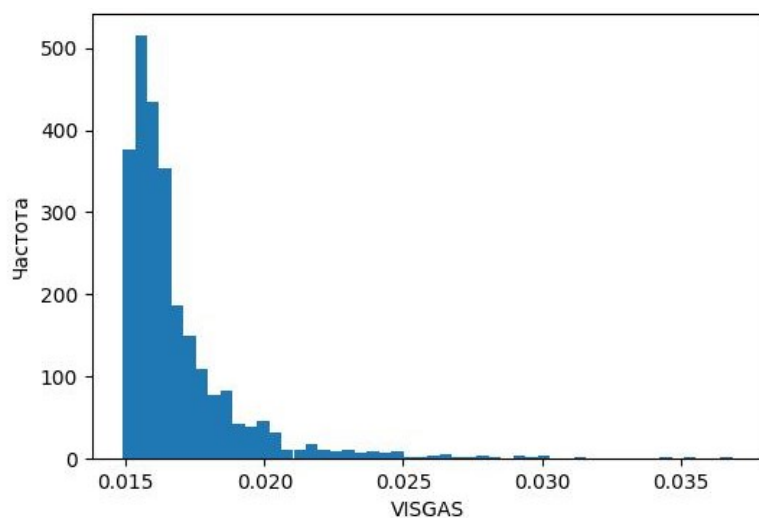
```
DGAS
count      2572.000000
mean        107.172287
std          35.191692
min          63.632900
25%          82.373425
50%          97.100100
75%         122.126700
max         319.476200
Name: DGAS, dtype: float64
Медиана: 97.1001
Дисперсия: 1238.4551775964776
Коэффициент вариации: 0.32836559548480504
```


10. VISLIQ - Вязкость жидкости



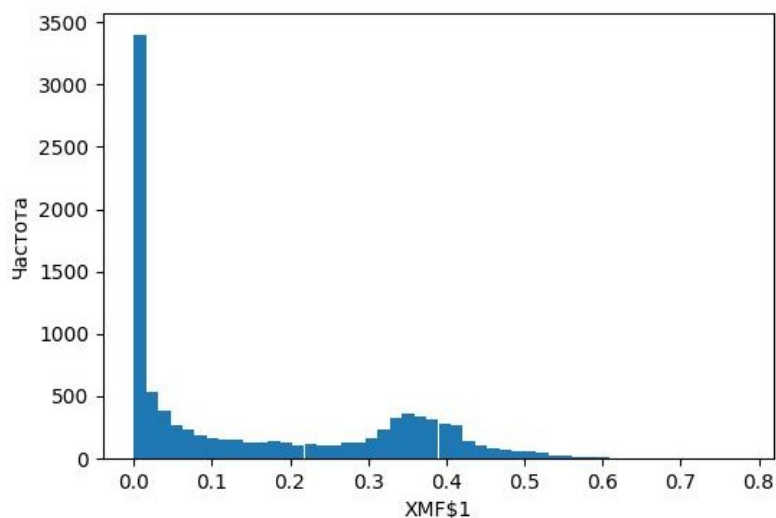
VISLIQ
count 9652.000000
mean 0.239078
std 0.177803
min 0.020430
25% 0.088425
50% 0.166250
75% 0.378900
max 0.637800
Name: VISLIQ, dtype: float64
Медиана: 0.16625
Дисперсия: 0.03161383001633561
Коэффициент вариации: 0.7437028575714999

11. VISGAS - Вязкость жидкости



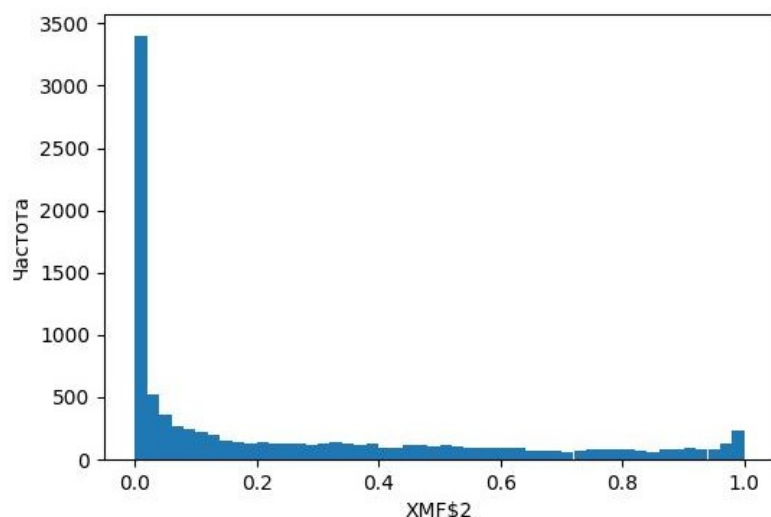
VISGAS
count 2572.000000
mean 0.016818
std 0.002164
min 0.014890
25% 0.015540
50% 0.016160
75% 0.017250
max 0.036880
Name: VISGAS, dtype: float64
Медиана: 0.01616
Дисперсия: 4.681500272388582e-06
Коэффициент вариации: 0.12864943919540744

12. XMF\$1 - Молярная концентрация компонента 1 в жидком состоянии



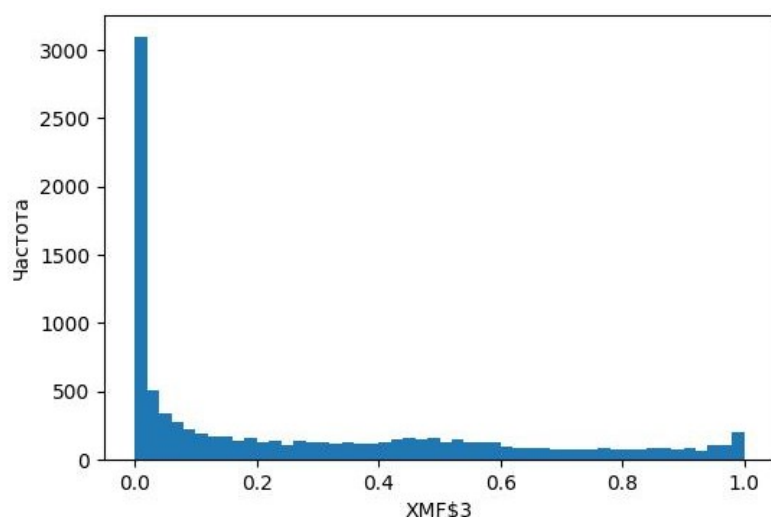
XMF\$1
count 9652.000000
mean 0.160512
std 0.170357
min 0.000000
25% 0.002663
50% 0.078029
75% 0.333554
max 0.780166
Name: XMF\$1, dtype: float64
Медиана: 0.0780292905
Дисперсия: 0.02902153617985588
Коэффициент вариации: 1.0613338781260164

13. XMF\$2 - Молярная концентрация компонента 2 в жидком состоянии



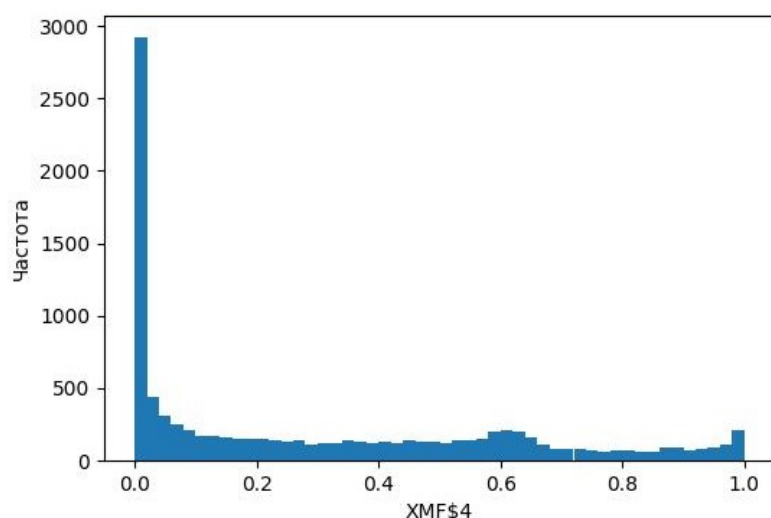
XMF\$2
count 9652.000000
mean 0.261190
std 0.312383
min 0.000000
25% 0.003529
50% 0.102800
75% 0.465752
max 0.999995
Name: XMF\$2, dtype: float64
Медиана: 0.102800167
Дисперсия: 0.09758312882020234
Коэффициент вариации: 1.1959986380307261

14. XMF\$3 - Молярная концентрация компонента 3 в жидком состоянии



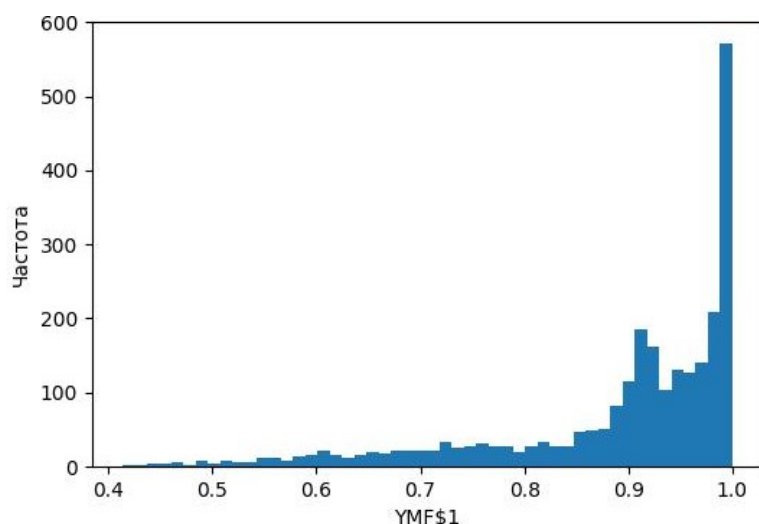
XMF\$3
count 9652.000000
mean 0.279997
std 0.310375
min 0.000000
25% 0.005504
50% 0.143455
75% 0.500806
max 0.999992
Name: XMF\$3, dtype: float64
Медиана: 0.143455357
Дисперсия: 0.09633278723731827
Коэффициент вариации: 1.108496265278872

15. XMF\$4 - Молярная концентрация компонента 4 в жидком состоянии



XMF\$4
count 9652.000000
mean 0.298301
std 0.311702
min 0.000000
25% 0.007659
50% 0.185055
75% 0.557264
max 0.999991
Name: XMF\$4, dtype: float64
Медиана: 0.18505541050000002
Дисперсия: 0.09715797102289259
Коэффициент вариации: 1.0449234436141281

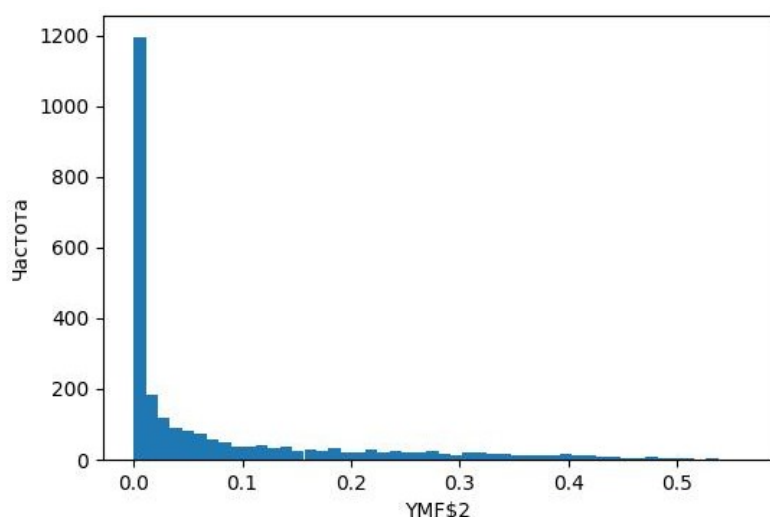
16. YMF\$1 - Молярная концентрация компонента 1 в газообразном состоянии



```

YMF$1
count      2572.000000
mean        0.892953
std         0.120905
min         0.413779
25%         0.858682
50%         0.929505
75%         0.984751
max         1.000000
Name: YMF$1, dtype: float64
Медиана: 0.929505409
Дисперсия: 0.014618107471184175
Коэффициент вариации: 0.1353995108513966
    
```

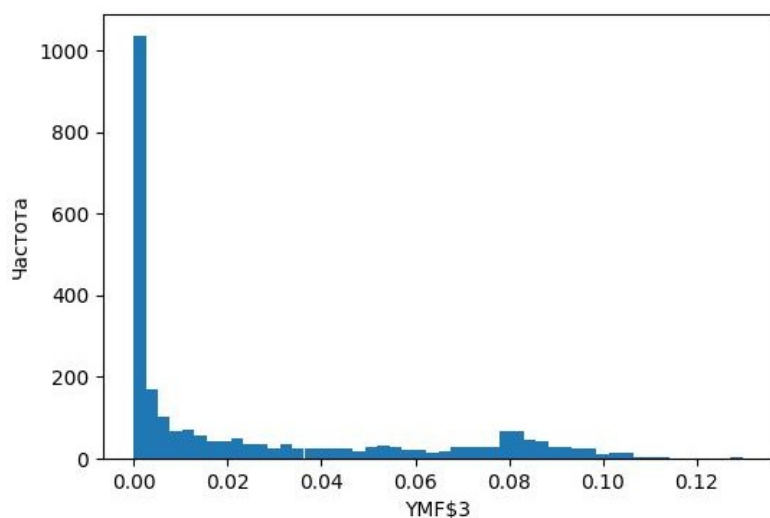
17. YMF\$2 - Молярная концентрация компонента 2 в газообразном состоянии



```

YMF$2
count      2572.000000
mean        0.079930
std         0.120822
min         0.000000
25%         0.000347
50%         0.016565
75%         0.112880
max         0.560332
Name: YMF$2, dtype: float64
Медиана: 0.016565194499999998
Дисперсия: 0.014597961238458723
Коэффициент вариации: 1.5115892734971743
    
```

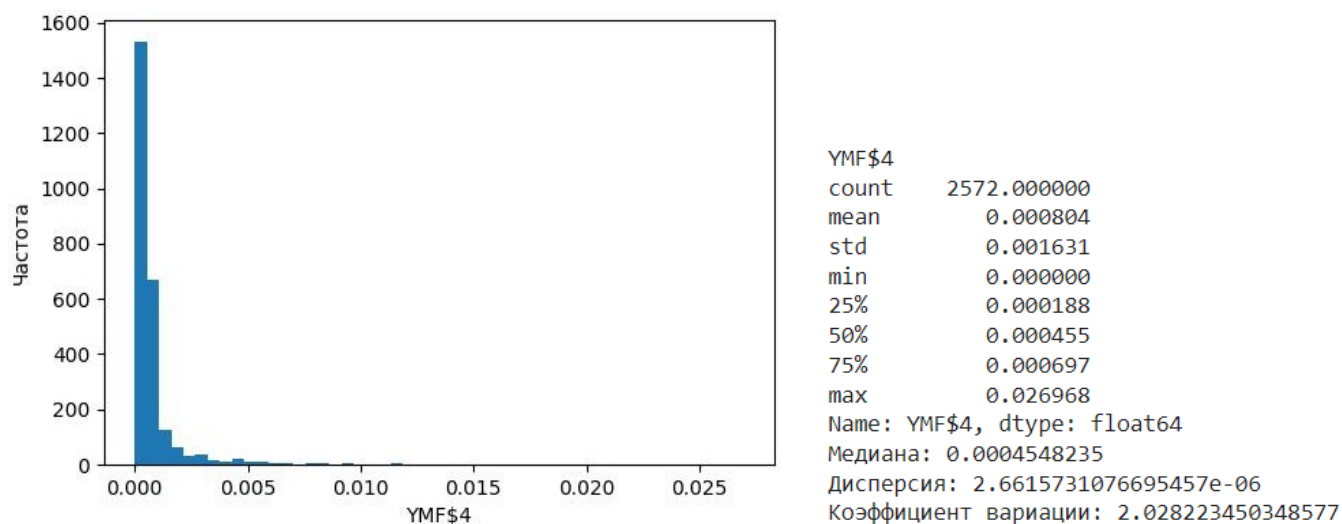
18. YMF\$3 - Молярная концентрация компонента 3 в газообразном состоянии



```

YMF$3
count      2572.000000
mean        0.026312
std         0.033177
min         0.000000
25%         0.000189
50%         0.007220
75%         0.050816
max         0.129821
Name: YMF$3, dtype: float64
Медиана: 0.0072198205
Дисперсия: 0.0011007449023887662
Коэффициент вариации: 1.2609052862523176
    
```

19. YMF\$4 - Молярная концентрация компонента 4 в газообразном состоянии



Корреляционная зависимость

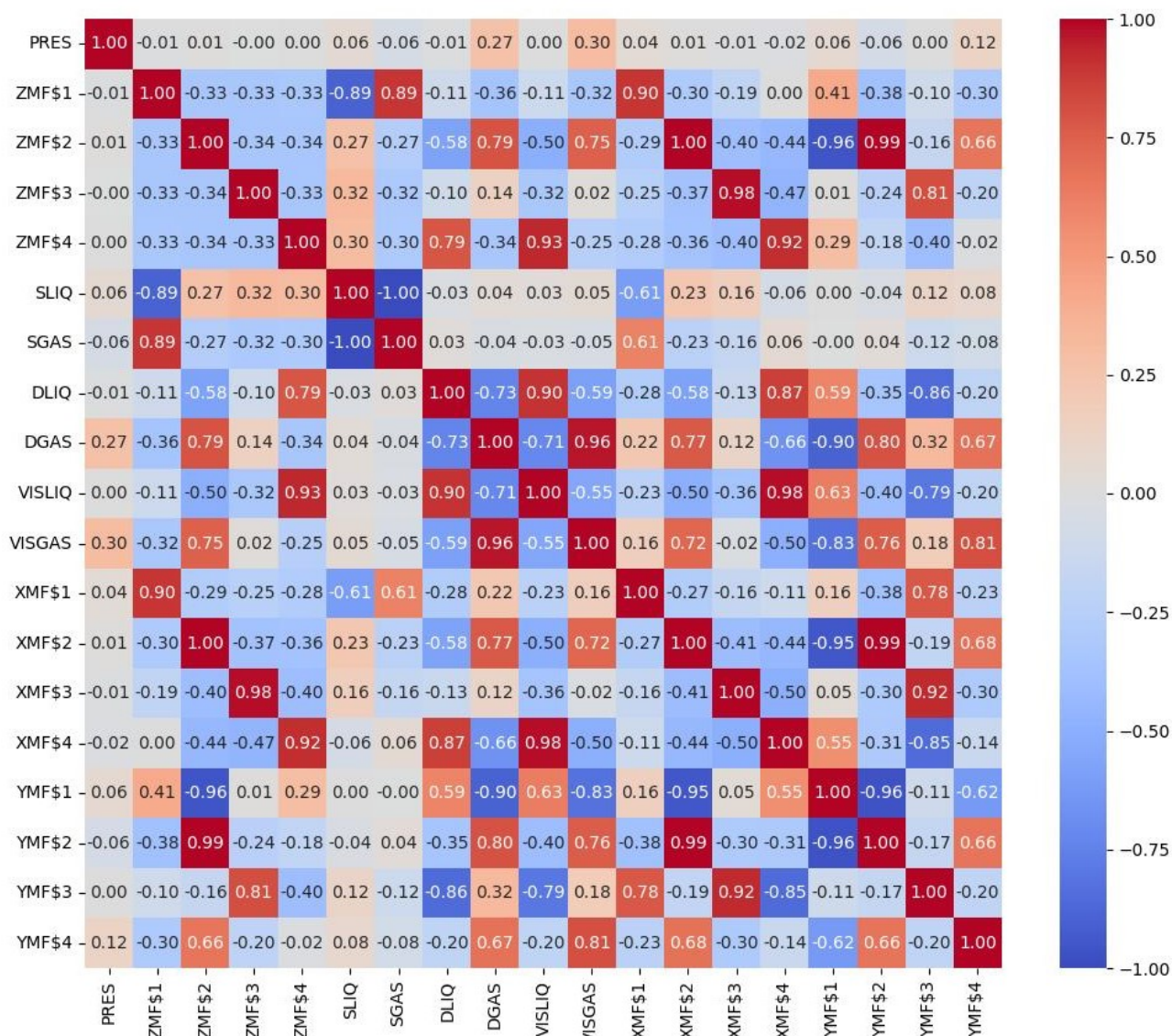


Рис. 2: Таблица корреляционной зависимости

Функциональная зависимость

$$XMF\$1 + XMF\$2 + XMF\$3 + XMF\$4 = 1,$$

$$YMF\$1 + YMF\$2 + YMF\$3 + YMF\$4 = 1,$$

$$ZMF\$1 + ZMF\$2 + ZMF\$3 + ZMF\$4 = 1.$$

Результаты поиска линейной зависимости:

- Между DLIQ и PRES нет линейной зависимости.
- Между DGAS и PRES есть очень незначительная линейная зависимость. R-squared принимает значение 0.081.
- XMF и PRES: Только первый компонент имеет линейную зависимость от давления, со значением R-squared 0.313.
- YMF и PRES: Ни у одного из компонентов нет линейной зависимости от PRES, для всех компонентов значение R-squared не превышает 0.015.
- SLIQ и SGAS - линейной зависимости с давлением PRES нет.