Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Уфимский государственный авиационный технический университет»

Кафедра вычислительной математики и кибернетики

Лабораторная работа №1

по дисциплине: «Компьютерная обработка экспериментальных данных»

«Корреляционный анализ»

Выполнили:

студенты группы МО-417

Ярыгин А.Э.

Алимгафаров А.Р.

Проверила:

Харисова Э. А.

Уфа 2022

**Цель:**

приобрести навыки корреляционного анализа.

**Задачи**

1. Выполнить задание для самостоятельной работы в соответствии с настоящим руководством по выполнению лабораторной работы;
2. Оформить отчет о выполнении лабораторной работы в соответствии с требованиями к его оформлению.

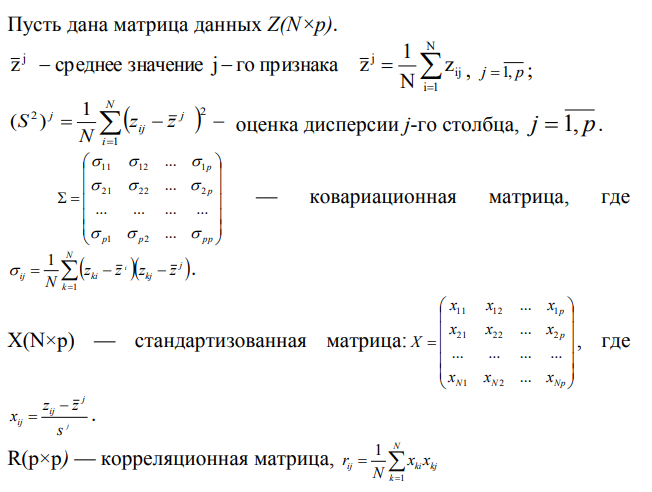
**Задание**

Пусть дана Z — матрица данных размером N×p.

1. Составить программу для вычисления
   1. средних по столбцам, дисперсий по столбцам;
   2. стандартизованной матрицы;
   3. ковариационной матрицы;
   4. корреляционной матрицы.
2. Проверить гипотезу о значимости коэффициентов корреляции между столбцами матрицы данных.

**Теоретическая часть**

1. **Нахождение корреляционной матрицы**



**Ковариация** (корреляционный момент, ковариационный момент) — в теории вероятностей и математической статистике мера линейной зависимости двух случайных величин.

**Ковариационная матрица** (или матрица ковариаций) в теории вероятностей — это матрица, составленная из попарных ковариаций элементов одного или двух случайных векторов.

**Корреляция (или корреляционная зависимость)** — статистическая взаимосвязь двух или более случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

**Корреляционная матрица (матрица корреляций)** – это квадратная таблица, заголовками строк и столбцов которой являются обрабатываемые переменные, а на пересечении строк и столбцов выводятся коэффициенты корреляции для соответствующей пары признаков.

**Стандартизированная матрица** – форма к которой принято приводить матрицу при плохой обособленности.

1. **Оценка значимости коэффициента корреляции**

Пусть имеются статистические гипотезы: H0: ρ(x, y)=0, связи между признаками x и y нет. H1: ρ(x, y)≠0, то есть связь есть. Здесь ρ(x, y)- коэффициент корреляции между x и y.

|  |  |  |
| --- | --- | --- |
| Действие  Состоя-  ние природы | H0 принимаем | H0 отвергаем |
| Верна H0 | Верное решение | α |
| Верна H1 | β | Верное решение |

α — вероятность ошибки первого рода — вероятность отвергнуть верную гипотезу, β — вероятность ошибки второго рода — вероятность принять неверную гипотезу. Надо сформулировать такое правило, чтобы α и β были достаточно малыми. В математической статистике показано, что статистика

при условии, что H0 справедлива, подчиняется закону распределения Стьюдента.

1. **Алгоритм проверки статистической гипотезы о**

**значимости коэффициента корреляции.**

1) Пусть имеются экспериментальные данные

(x1, y1)

(x2, y2)

...

(xn, yn).

Вычисляем r(x, y) — выборочный коэффициент корреляции.

2) Задаемся приемлемой для нас вероятностью ошибки α, пусть

α=0,05.

3) Вычисляем статистику t.

4) По выбранному α и числу степеней свободы f=n-2 определяем

tтабличное.

5) Правило вынесения решения: если |tрасч| ≥ tтабл, то справедлива

гипотеза H1, в противном случае — H0, а отличие от нуля

коэффициента корреляции обусловлено случайными причинами.

**Ход работы**

1. В качестве матрицы Z возьмём матрицу значений различный признаков для областей РФ. Матрица Z будет содержать 75 строк и 10 столбцов. В качестве инструмента разработки воспользуемся языком python. Данные, которые необходимо найти в ходе лабораторной работы будут записаны в файлы excel.

Загрузим исходные данные:



Данные можно рассмотреть на рисунке 2:

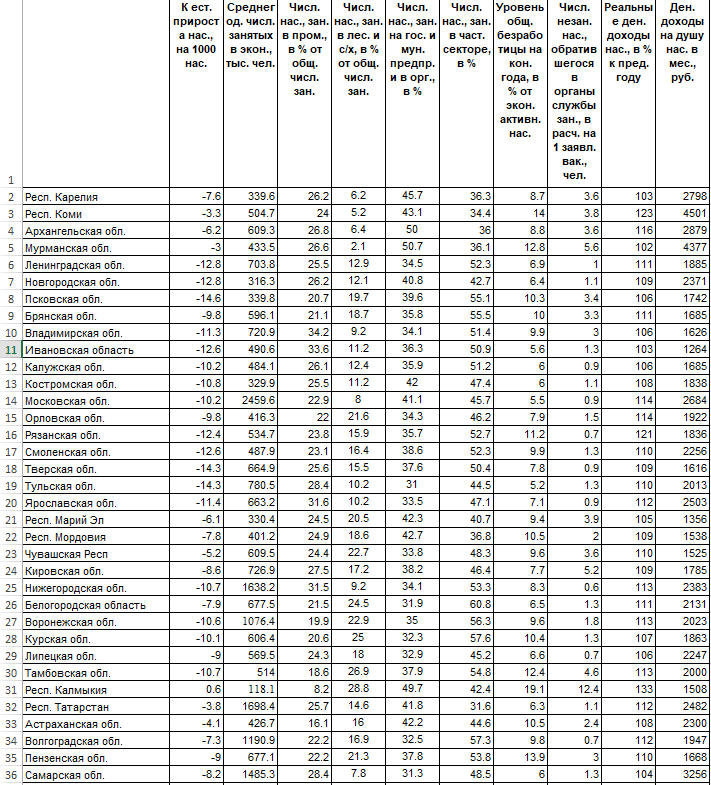


Рисунок 1 - Входные данные (1 часть).

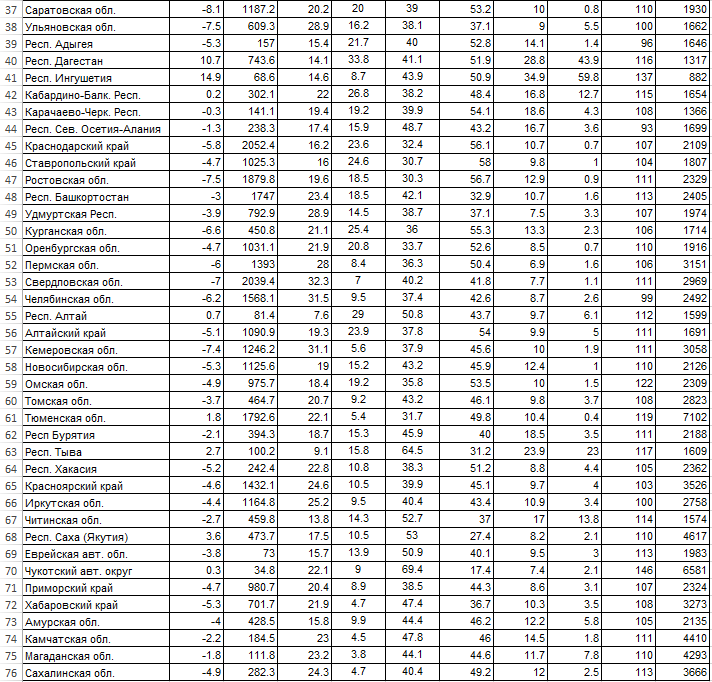
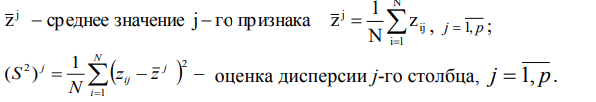


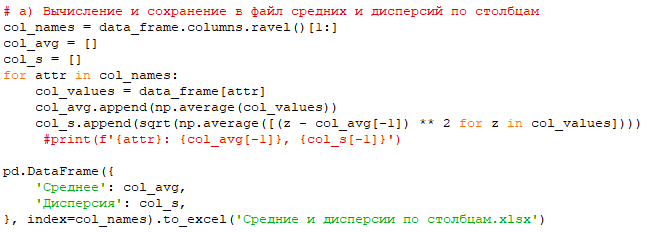
Рисунок 1 - Входные данные (часть 2).

1. В первую очередь найдём и сохраним в файл значения средних по столбцам и дисперсий по столбцам.

Для этого воспользуемся следующими формулами:



Код для программного выполнения описанных выше действий:

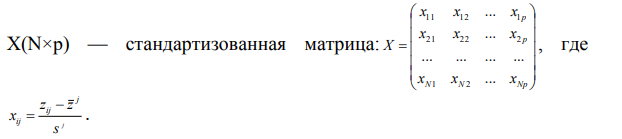


Результат можно рассмотреть на рис. 2



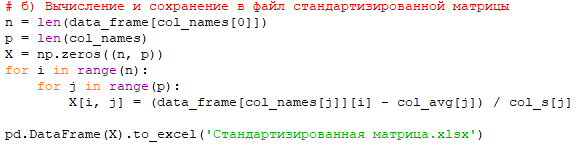
Рисунок 2 – Значения средних и дисперсий по столбцам.

1. Воспользовавшись следующей формулой:



на основе вычисленных нами ранее данных найдём стандартизированную матрицу и сохраним её в файл.

Код для программного выполнения описанных выше действий:



Результат можно рассмотреть на рисунке 3:

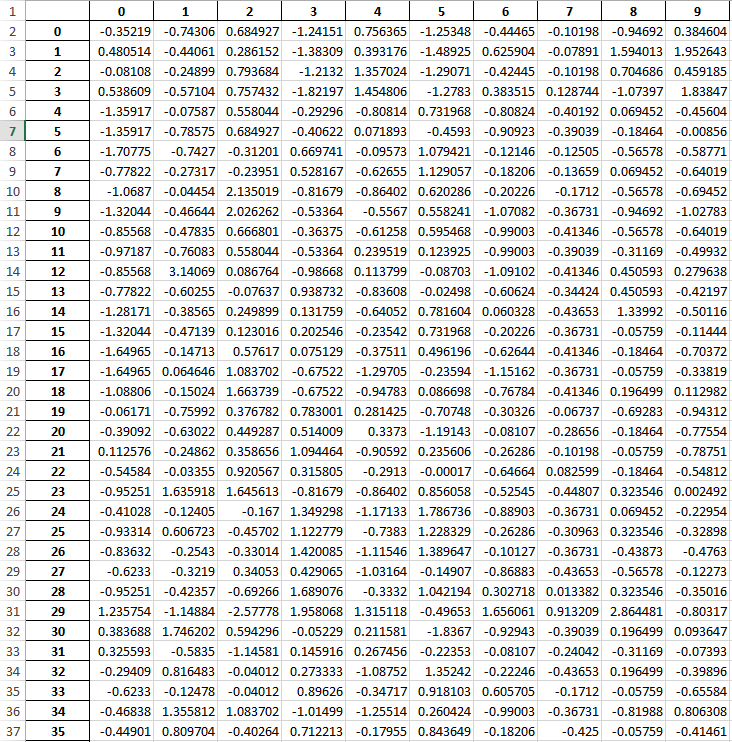


Рисунок 3 – Стандартизированная матрица (часть 1).

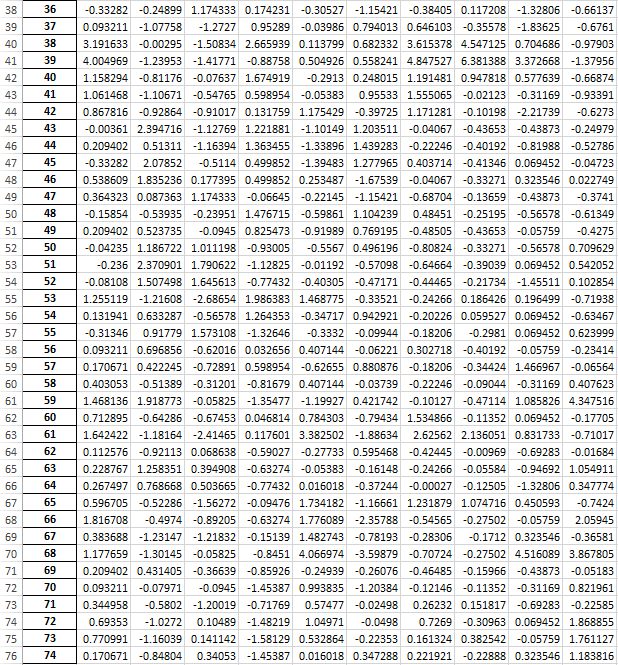
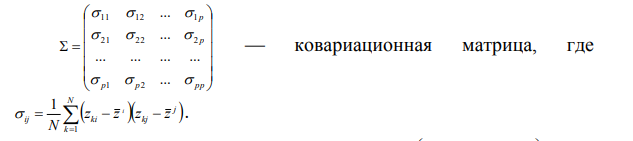
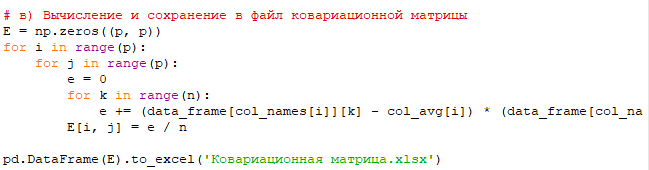


Рисунок 3 – Стандартизированная матрица (часть 2).

1. Найдём матрицу ковариации и сохраним её в файл, воспользовавшись вычисленными ранее величинами и следующей формулой:



Код для программного выполнения описанных выше действий:



Результат можно рассмотреть на рисунке 4:

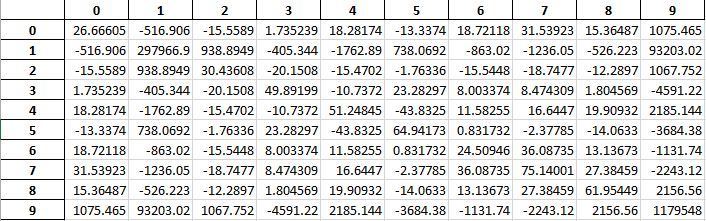
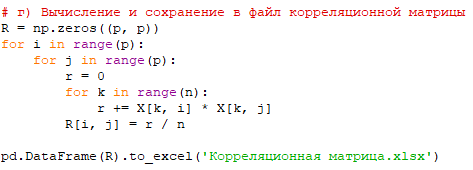


Рисунок 4 – Ковариационная матрица.

1. Найдём корреляционную матрицу, воспользовавшись вычисленными ранее величинами и следующей формулой:



Код для программного выполнения описанных выше действий:



Результат можно рассмотреть на рисунке 5:

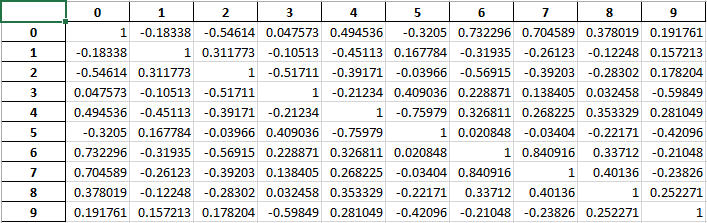


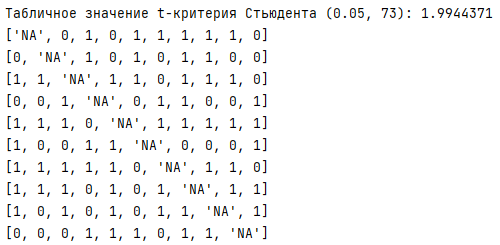
Рисунок 5 – Корреляционная матрица.

1. Проверим гипотезу о значимости коэффициентов корреляции между столбцами матрицы данных. Для этого нам нужно будет воспользовавшись таблицей значений t-критерия Стьюдента найти значение tтабл . Затем вычислить tрасч по формуле для каждого элемента корреляционной матрицы:

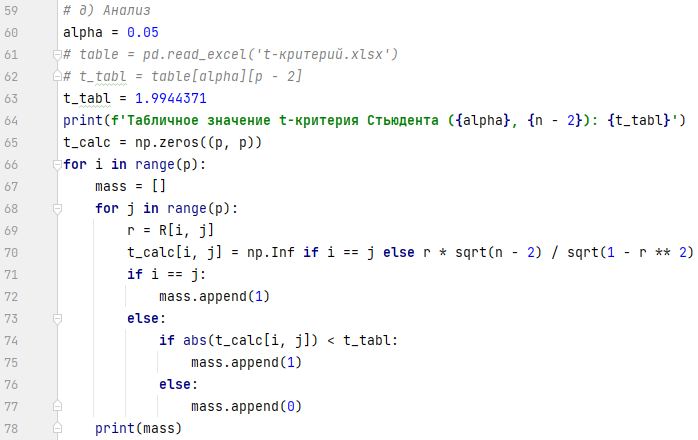
Матрица корреляции была рассчитана шагом ранее, поэтому теперь всё, что нам нужно сделать - расcчитать tрасч, подставив вычисленные ранее значения. Далее, сохраним матрицу A(10\*10), где если между признаками i и j

есть связь, то a[i][j]=1, иначе a[i][j]=0.

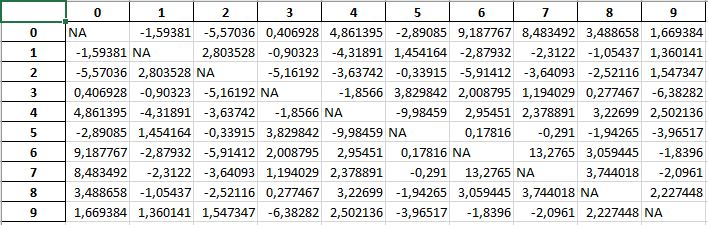
Вывод матрицы A на экран:

  
Рисунок 6 – Матрица А

Код:

  
Рисунок 7 – Код анализа

Вывод матрицы статистик:

  
Рисунок 8 – Матрицa статистик

**Вывод**

В ходе выполнения лабораторной работы мы реализовали программу для выполнения корреляционного анализа.