МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«Кубанский государственный университет»**

**(ФГБОУ ВО «КубГУ»)**

**Кафедра вычислительных технологий**

**ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №8**

**Дисциплина: Обработка больших данных**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Минин К.С.

(подпись)

Факультет Компьютерных технологий и прикладной математики

Направление подготовки02.03.02 Фундаментальная информатика и информационные технологии курс 3

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Приходько Т.А.

Краснодар 2022

**Лабораторная работа № 8**

**«Корреляционный и регрессионный анализ данных»**

**Цель работы:**

Ознакомиться с понятием корреляционный и регрессионный анализ данных, некоторыми функциями языка R, осуществляющими этот вид анализа, принципами их работы. Научиться оценивать связь между переменными и оценивать степень этой связи.

**Описание данных:** различные экономические показатели стран за период 1989-2017 (ВВП, Процентный рост ВВП, Процент рождения скилованным персоналом, Рождаемость, Национальный доход, Процент безработных и т.п). Обработанный датасет для Китая представлен на рисунке 1.

**Ход работы:**

Классическим инструментом для измерения лилейной зависимости между двумя наборами данных является коэффициент корреляции. Коэффици­ент корреляции — это числовая величина, находящаяся в интервале от -1 до +1. Чем она больше по модулю (т.е. ближе к 1 или к -1), тем выше линейная связь между наборами данных. Знак коэффициента корреляции показывает, в одном ли направлении изменяются наборы данных. Если один из наборов возрастает, а второй убывает, то коэффициент корреляции отрицателен, а если оба набора од­новременно возрастают или убывают, то коэффициент корреляции положителен. Значение коэффициента корреляции по модулю равное 1 соответствует точной линейной зависимости между двумя наборами данных. Линейная зависимость является самой любимой у экспериментаторов всех мастей.

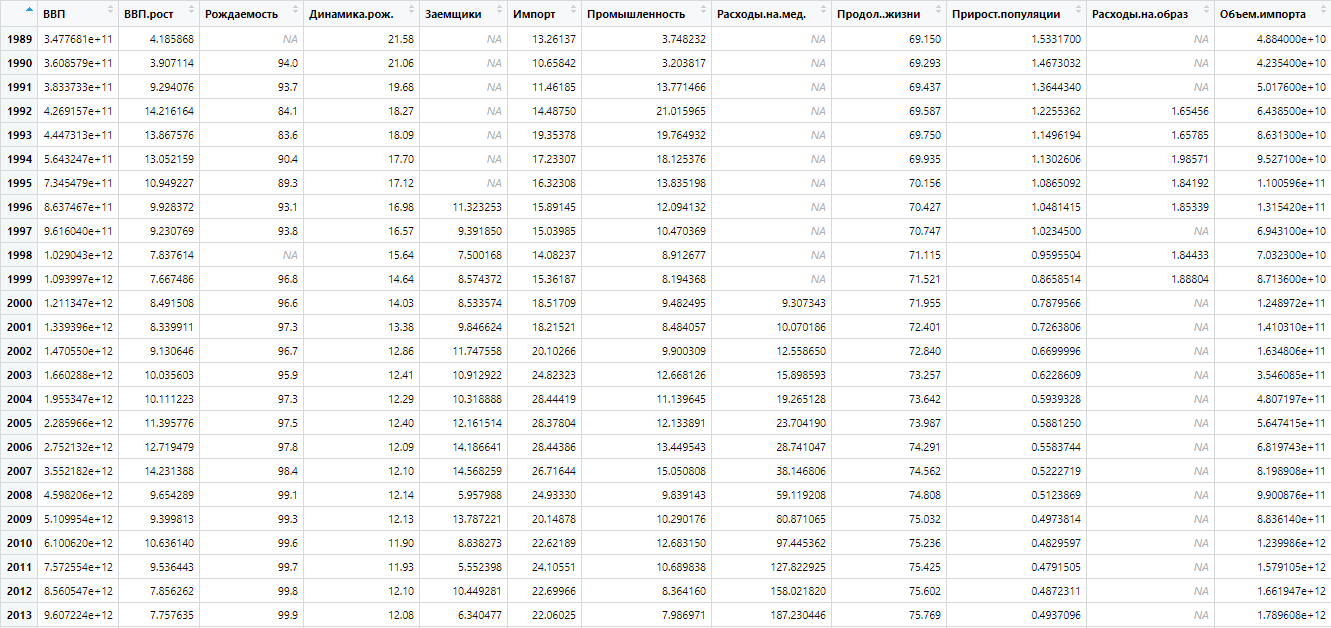


Рисунок 1 – Фрагмент исходных данных

Построим график прироста ВВП за весь период на рисунке 2.

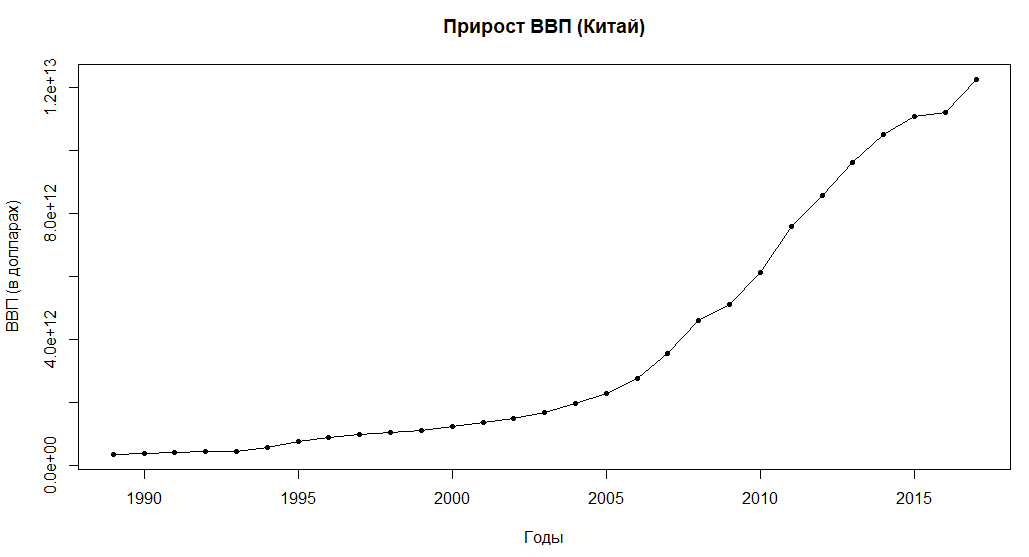
****

Рисунок 2 – График роста ВВП в Китае.

Вычислим корреляционную матрицу для каждого показателя и построим их визуализацию (рисунок 3).

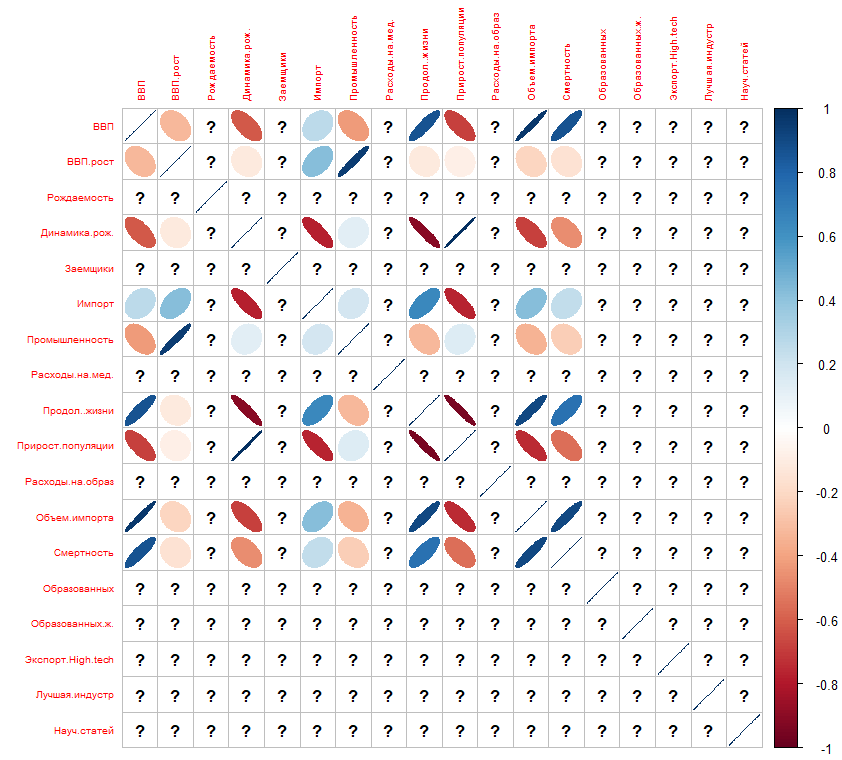
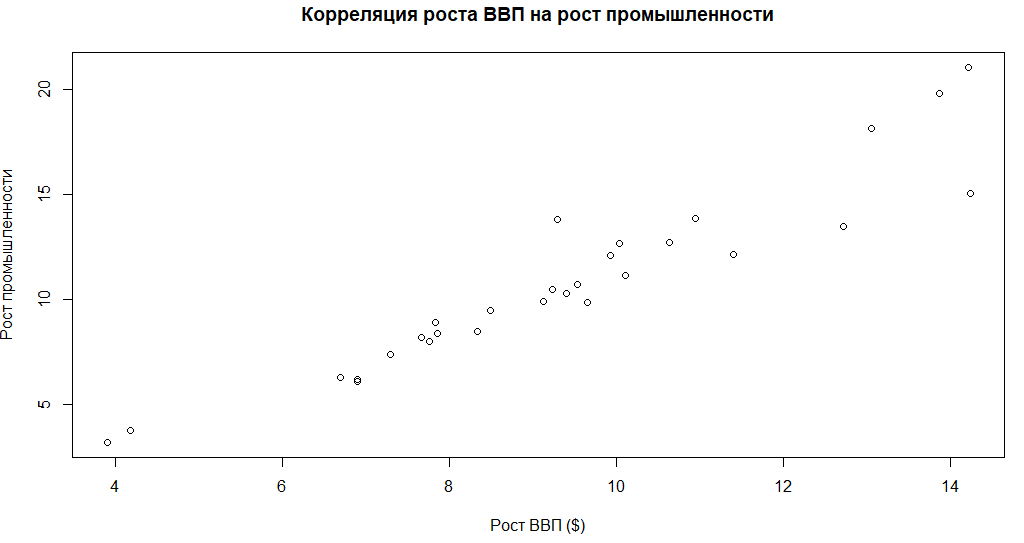
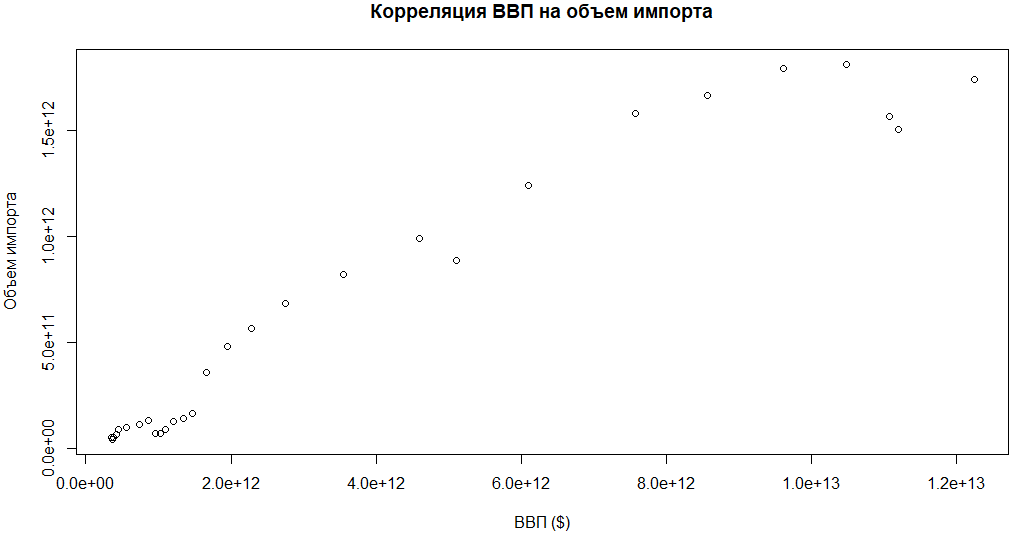
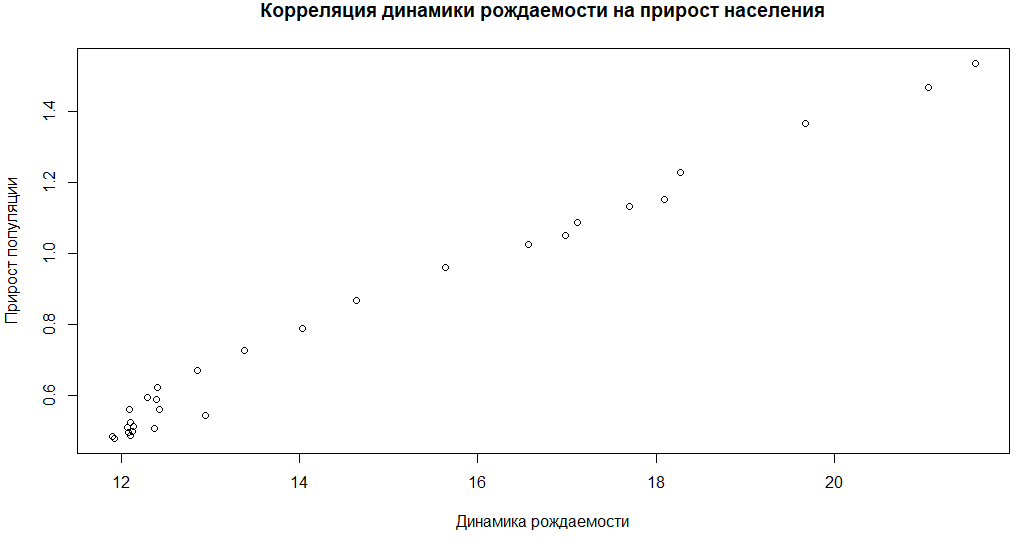
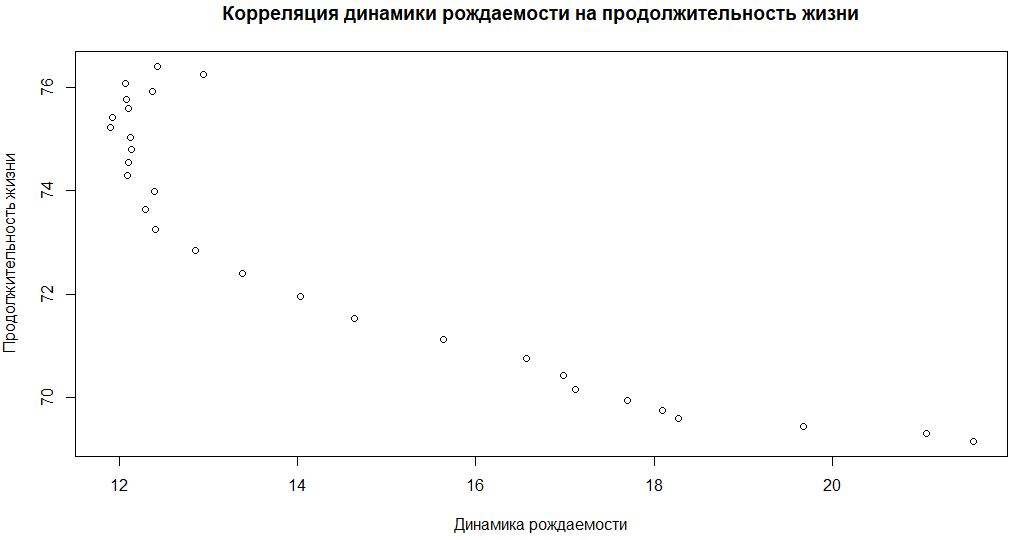
****

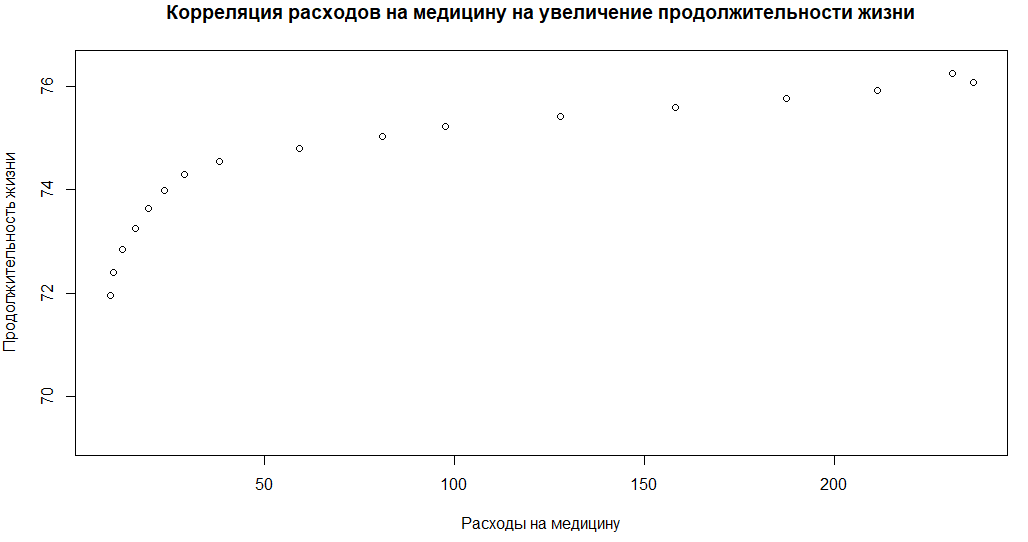
Рис. 3 – Корреляционная матрица.

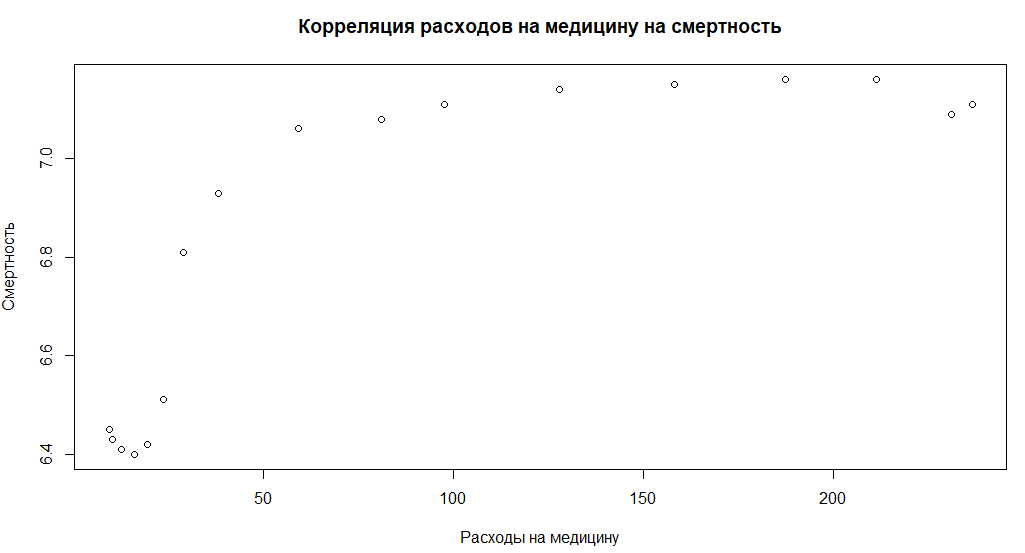












Вывод: Ознакомился с некоторыми статистическими тестами, принципами их работы. Научился оценивать нормальность распределения выборки, а также выполнил оценку статистических гипотез.