Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**Отчёт**

**по лабораторной работе №8**

**Дисциплина: ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

**Тема: «Проверка статистических гипотез»**

Работу выполнила \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ К.В.Стасюк

Направление подготовки 02.03.02 Фундаментальная информатика и

информационные технологии

Направленность (профиль) Математическое и программное обеспечение

компьютерных технологий

Преподаватель

ассистент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.И.Шиян

**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.

**Задание:**

1. С указанного ниже ресурса (Всемирный Банк - World Development Indicators database) сформирована CSV-таблица, содержащая информацию по 18 странам мира и по 22 характеристикам по каждой стране за годы(1989-2018).
2. Необходимо выбрать данные по одной стране, согласно варианту, построить кривую прироста ВВП, проанализировать взаимное влияние атрибутов, отобразить корреляцию:
   1. Роста ВВП и прироста населения.
   2. Прироста населения на динамику безработицы.
   3. Изменения расходов на медицину и увеличения продолжительности жизни и смертность.
   4. Прирост людей с высшим образованием на рост экспорта товаров и на прирост высокотехнологичного производства.
   5. Расходов на образование на – кумулятивный прирост бакалавров среди женщин.
   6. Прирост людей с высшим образованием на развитие высоких технологий (прирост статей в научных журналах).
   7. С помощью регрессионного анализа найдите зависимые переменные и поясните влияние на них независимых переменных.
   8. С помощью функции predict() построить прогноз по любому понравившемуся атрибуту.

Данные NA не учитывать. Вариант: Беларусь.

**Ход работы:**

Из исходного csv файла данные были импортированы в RStudio. На рисунке 1 представлен фрагмент таблицы.

**Изображение выглядит как текст, число, Шрифт, снимок экрана

Автоматически созданное описание**

Рисунок 1 – Фрагмент исходной таблицы

Была произведена выборка данных для Беларуси и запись их в новую таблицу, с заменой отсутствующего значения на значение NA для удобства работы.

Изображение выглядит как текст, число, документ, Шрифт

Автоматически созданное описание

Рисунок 2 – Фрагмент таблицы с данными о Беларуси

Построим график роста ВВП Франции на протяжении 30 лет. График изображён на рисунке 3.

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

Рисунок 3 – Рост ВВП Беларуси на протяжении 25 лет

Отобразим корреляцию роста ВВП и прироста населения. Рассмотрим корреляцию 2-мя методами: Спирмена и Пирсона.

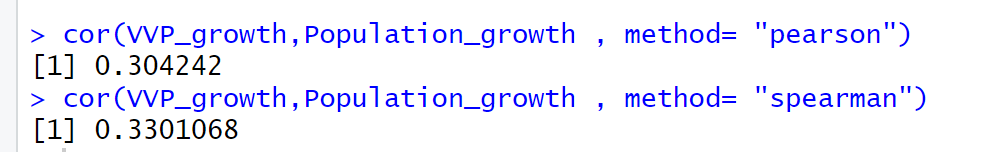


Рисунок 4 – Корреляция роста ВВП и прироста населения Беларуси

По полученному значению видно, что корреляция между ростом ВВП и приростом населения слабая положительная.

Рассмотрим корреляцию между изменением расходов на медицину и увеличением продолжительности жизни. На рисунке 5 представлена корреляция по 2-м методам. Корреляция является умеренной положительной.

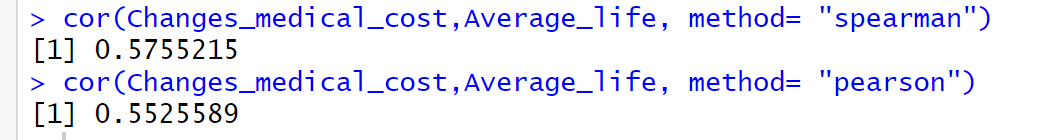


Рисунок 5 – Корреляция между изменением расходов на медицину и увеличением продолжительности жизни

Теперь рассмотрим между изменением расходов на медицину и смертностью. Данные значения предполагают слабую отрицательную корреляцию, указывая на то, что по мере увеличения изменений в расходах на медицинское обслуживание наблюдается тенденция к незначительному снижению уровня смертности

Изображение выглядит как текст, Шрифт, снимок экрана, линия

Автоматически созданное описание

Рисунок 7 – Корреляция между изменением расходов на медицину и смертностью

Изобразим визуально результаты исследования корреляции между 3-мя характеристиками. Значения коэффициентов корреляции рисуются в виде эллипсов, отражающих форму плотности двумерного нормального распределения с данным значением корреляции между компонентами. Чем ближе значение коэффициента корреляции к +1 или -1 — тем более вытянутым становится эллипс. Наклон эллипса отражает знак.

Изображение выглядит как текст, снимок экрана, диаграмма, дизайн

Автоматически созданное описание

Рисунок 8 – Визуальное представление результатов исследования корреляции

Провести аналогичные действия

* для прироста людей с высшим образованием на рост экспорта товаров и на прирост высокотехнологичного производства
* для расходов на образование на кумулятивный прирост бакалавров среди женщин
* для прироста людей с высшим образованием на развитие высоких технологий (прирост статей в научных журналах)

невозможно, так как одна из метрик содержит 1 или 0 значений.

Разберёмся, на сколько ВВП зависит от годового прироста населения, экспорта товаров и технического роста производства. На рисунке 9 изображены коэффициенты зависимости.

Изображение выглядит как текст, снимок экрана, Шрифт, алгебра

Автоматически созданное описание

Рисунок 9 – Коэффициенты зависимости

Составим матрицу рассеивания, изображённую на рисунке 10.

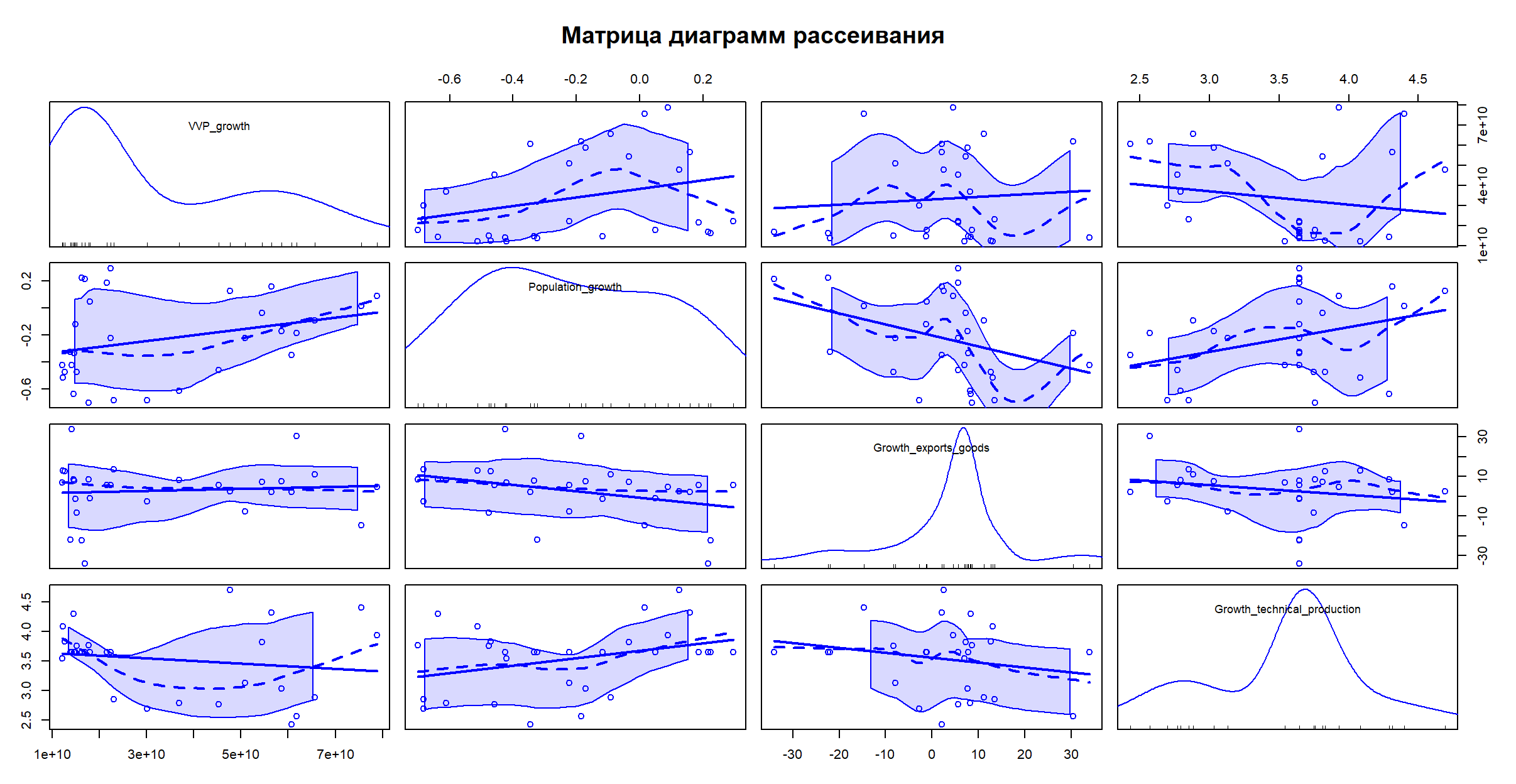


Рисунок 10 – Матрица диаграмм рассеивания