Министерство науки и высшего образования РФ

Федеральное государственное бюджетное

образовательное учреждение высшего образования

ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет автоматики и вычислительной техники

Отчет по лабораторной работе №0

**Исследование методов прогнозирования с помощью среды R. Установка и настройка пакетов**

дисциплина «Теория и применение методов прогнозирования»

|  |  |
| --- | --- |
| Выполнил: студент группы ИВТм-1301 | \_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Савин Д.А. / |
|  |  |
| Проверил: к.т.н., доцент кафедры ЭВМ | \_\_\_\_\_\_\_\_\_\_\_\_\_\_/ Коржавина А.С. / |

Киров 2022

##### Задание

1. Выполнить описанные выше действия с тестовым набором данных Dreissena.

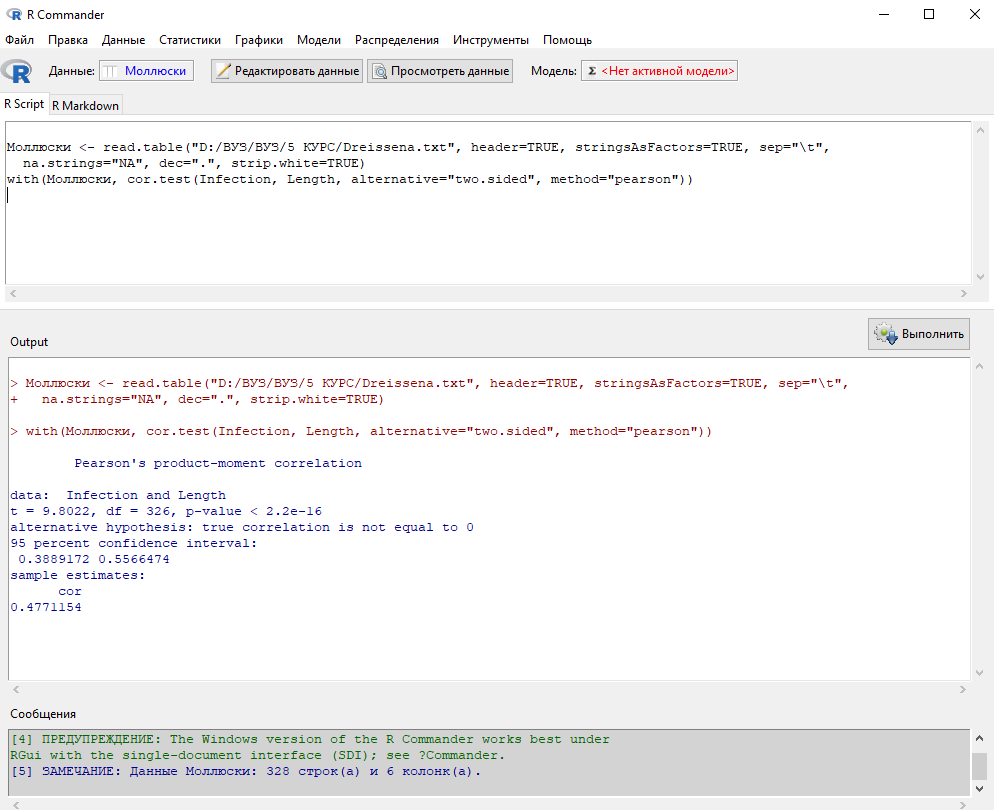


Рисунок 1 – Корреляционный тест

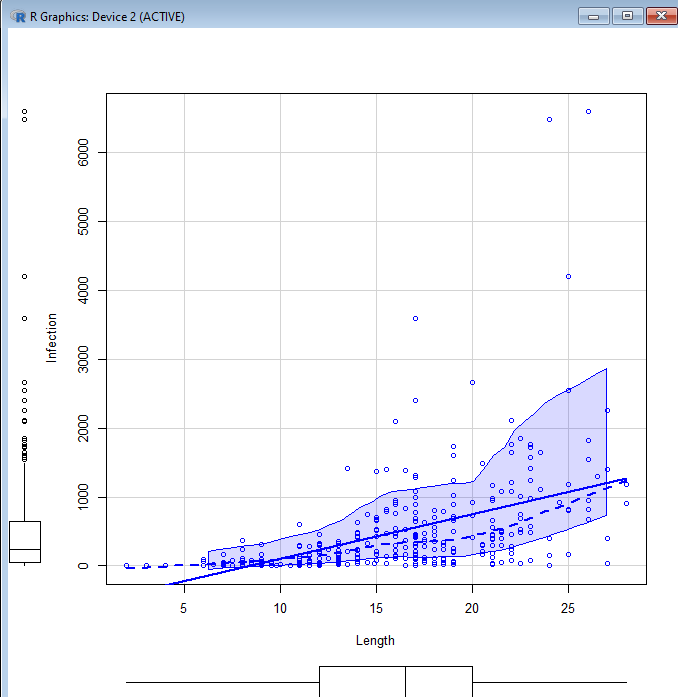


Рисунок 2 – Точечный график

1. Выполнить описанные выше действия с тестовым набором данных nybirths.dat.

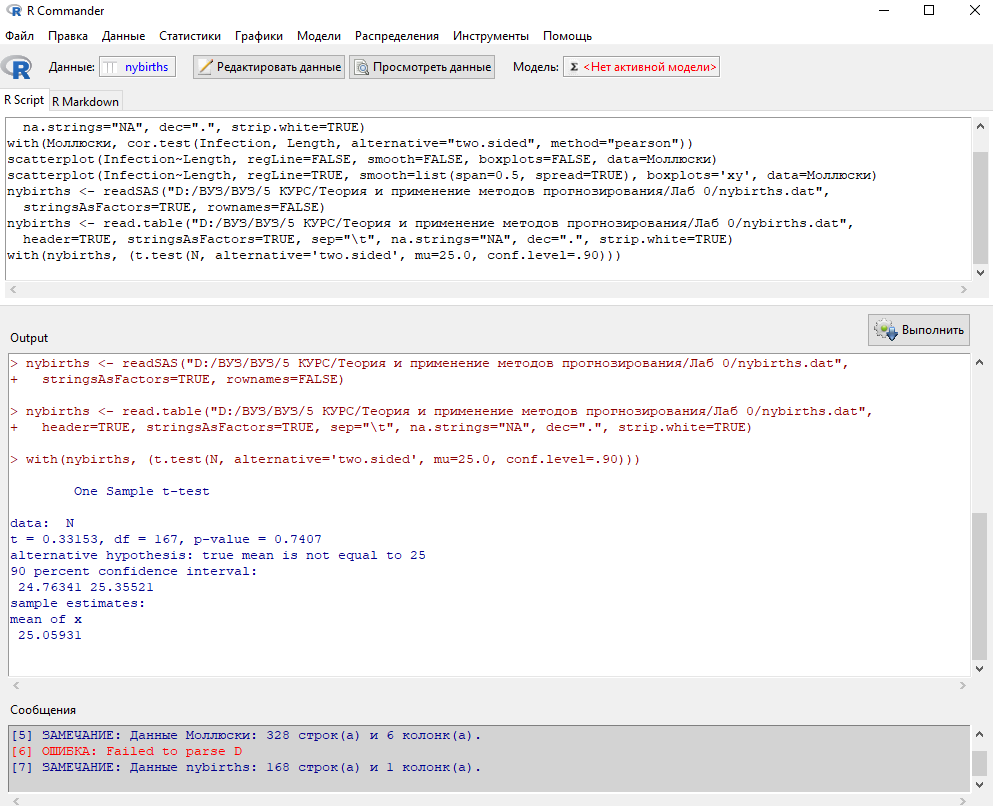


Рисунок 3 – Одновыборочный t-тест

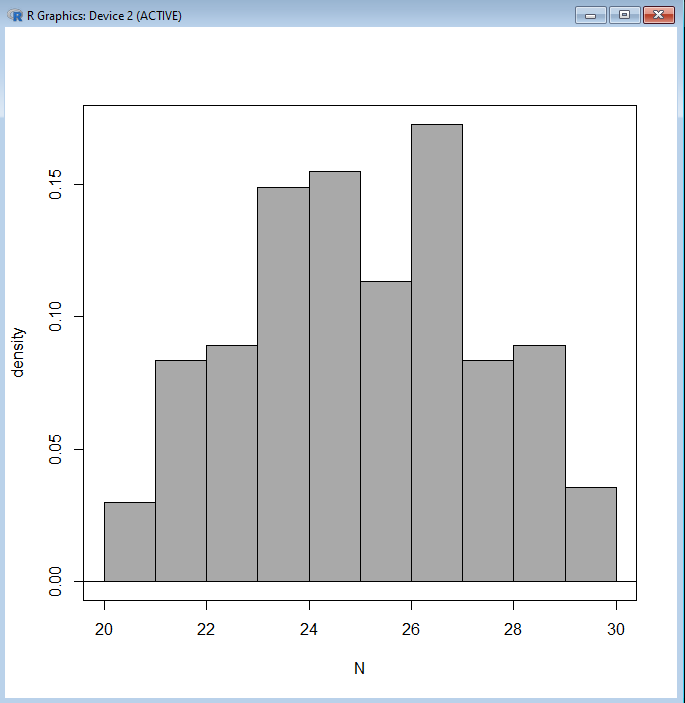


Рисунок 4 – Гистограмма

1. Изучить основные функции пакета R Commander и определить, какие из них могут быть полезны для решения реальной прикладной задачи.

|  |
| --- |
| which.max(x) — индекс элемента с максимальным значением  which.min(x) — индекс элемента с минимальным значением  rev(x) — реверсирует порядок элементов  sort(x) — сортирует элементы объекта по возрастанию  cut(x,breaks) — делит вектор на равные интервалы  match(x, y) — ищет элементы x, которые есть в y  which(x == a) — возвращает порядковые элементы x, которые равны a  na.omit(x) — исключает отсутствующие значения объекта  na.fail(x) — бросает исключение, если объект содержит отсутствующие значения  unique(x) — исключает из объекта повторяющиеся элементы  table(x) — создаёт таблицу с количеством повторений каждого уникального элемента  subset(x, …) — возвращает подмножество элемента, которое соответствует заданному условию  sample(x, size) — возвращает случайный набор размера size из элементов x  replace(x, list, values) — заменяет значения x c индексами из list значениями из values  append(x, values) — добавляет элементы values в вектор x  plot(x) — график x  plot(x, y) — график зависимости y от x  hist(x) — гистограмма  barplot(x) — столбчатая диаграмма  dotchart(x) — диаграмма Кливленда  pie(x) — круговая диаграмма  boxplot(x) — график типа “коробочки с усами”  sunflowerplot(x, y) — то же, что и plot(), однако точки с одинаковыми координатами изображаются в виде “ромашек”, количество лепестков у которых пропорционально количеству таких точек  coplot(x˜y | z) — график зависимости y от x для каждого интервала значений z  interaction.plot(f1, f2, y) — если f1 и f2 — факторы, эта фукнция создаст график со средними значениями y в соответствии со значениями f1 (по оси х) и f2 (по оси у, разные кривые)  matplot(x, y) — график зависимости столбцов y от столбцов x  fourfoldplot(x) — изображает (в виде частей окружности) связь между двумя бинарными переменными в разных совокупностях  assocplot(x) — график Кохена-Френдли  mosaicplot(x) — мозаичный график остатков лог-линейной регрессии  pairs(x) — если х - матрица или таблица данных, эта функция изобразит диаграммы рассеяния для всех возможных пар переменных из х  plot.ts(x), ts.plot(x) — изображает временной ряд  qqnorm(x) — квантили  qqplot(x, y) — график зависимости квантилей y от квантилей х  contour(x, y, z) — выполняет интерполяцию данных и создает контурный график  filled.contour(x, y, z) — то же, что contour(), но заполняет области между контурами определёнными цветами  image(x, y, z) — изображает исходные данные в виде квадратов, цвет которых определяется значениями х и у  persp(x, y, z) — то же, что и image(), но в виде трехмерного графика  stars(x) — если x — матрица или таблица данных, изображает график в виде “звезд” так, что каждая строка представлена “звездой”, а столбцы задают длину сегментов этих “звезд”  symbols(x, y, …) — изображает различные символы в соответствии с координатами  termplot(mod.obj) — изображает частные эффекты переменных из регрессионной модели  points(x, y) — рисование точек  lines(x, y) — рисование линии  text(x, y, labels, …) — добавление текстовой надписи  mtext(text, side=3, line=0, …) — добавление текстовой надписи  segments(x0, y0, x1, y1) — рисование отрезка  arrows(x0, y0, x1, y1, angle= 30, code=2) — рисование стрелочки  abline(a,b) — рисование наклонной прямой  abline(h=y) — рисование вертикальной прямой  abline(v=x) — рисование горизонтальной прямой  abline(lm.obj) — рисование регрессионной прямой  rect(x1, y1, x2, y2) — рисование прямоугольника  polygon(x, y) — рисование многоугольника  legend(x, y, legend) — добавление легенды  title() — добавление заголовка  axis(side, vect) — добавление осей  rug(x) — рисование засечек на оси X  locator(n, type = “n”, …) — возвращает координаты на графике, в которые кликнул пользователь  sd(x) — стандартное отклонение  var(x) — дисперсия  cor(x) — корреляционная матрица  var(x, y) — ковариация между x и y  cor(x, y) — линейная корреляция между x и y  aov(formula) — дисперсионный анализ  anova(fit,…) — дисперсионный анализ для подогнанных моделей fit  density(x) — ядерные плотности вероятностей  binom.test() — точный тест простой гипотезы о вероятности успеха в испытаниях Бернулли  pairwise.t.test() — попарные сравнения нескольки независимых или зависимых выборок  prop.test() — проверка гипотезы о том, что частоты какого-либо признака равны во всех анализируемых группах  t.test() — тест Стьюдента  rnorm(n, mean=0, sd=1) — нормальное распределение  rexp(n, rate=1) — экспоненциальное распределение  rgamma(n, shape, scale=1) — гамма-распределение  rpois(n, lambda) — распределение Пуассона  rweibull(n, shape, scale=1) — распределение Вейбулла  rcauchy(n, location=0, scale=1) — распределение Коши  rbeta(n, shape1, shape2) — бета-распределение  rt(n, df) — распределение Стьюдента  rf(n, df1, df2) — распределение Фишера  rchisq(n, df) — распределение Пирсона  rbinom(n, size, prob) — биномиальное распределение  rgeom(n, prob) — геометрическое распределение  rhyper(nn, m, n, k) — гипергеометрическое распределение  rlogis(n, location=0, scale=1) — логистическое распределение  rlnorm(n, meanlog=0, sdlog=1) — логнормальное распределение  rnbinom(n, size, prob) — отрицательное биномиальное распределение  runif(n, min=0, max=1) — равномерное распределение  sin(x), cos(x), tan(x), asin(x), acos(x), atan(x), atan2(y, x), log(x), log(x, base), log10(x), exp(x) — элементарные математические функции  min(x), max(x) — минимальный и максимальный элементы объекта  range(x) — вектор из минимального и максимального элемента объекта  pmin(x, y), pmax(x, y) — возвращают вектор с минимальными (максимальными) для каждой пары x[i], y[i]  sum(x) — сумма элементов объекта  prod(x) — произведение элементов объекта  diff(x) — возвращает вектор из разниц между соседними элементами  mean(x) — среднее арифметическое элементов объекта  median(x) — медиана (средний элемент) объекта  weighted.mean(x, w) — средневзвешенное объекта x (w определяет веса)  round(x, n) — округляет x до n знаков после запятой  cumsum(x), cumprod(x), cummin(x), cummax(x) — кумулятивные суммы, произведения, минимумы и максимумы вектора x (i-ый элемент содержит статистику по элементам x[1:i])  union(x, y), intersect(x, y), setdiff(x,y), setequal(x,y), is.element(el,set) — операции над множествами: объединение, пересечение, разность, сравнение, принадлежность  Re(x), Im(x), Mod(x), Arg(x), Conj(x) — операции над комплексными числами: целая часть, мнимая часть, модуль, аргумент, сопряжённое число  fft(x), mvfft(x) — быстрое преобразование Фурье  choose(n, k) — количество сочетаний  rank(x) — ранжирует элементы объекта |

1. Подготовить фрагмент данных из реальной задачи.

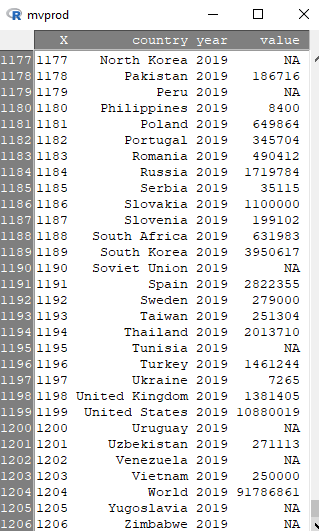


Рисунок 5 – Производство автомобилей по странам, 1950–2019 гг.

1. Выполнить базовые функции для выбранного фрагмента данных.

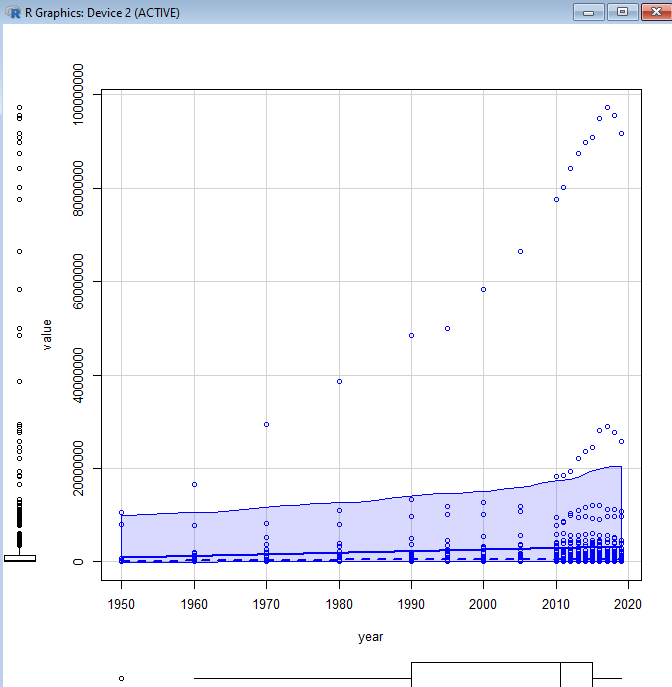


Рисунок 6 – Увеличение производства автомобилей со временем.

**Вывод**

Среда R является мощным инструментом для обработки массивов данных, а также проведения анализа над ними, включая графическое отображение. В качестве функций для проведения анализа можно воспользоваться такими функциями:

1. Графики – гистограмма, точечный график, линейный график.
2. Статистики – корреляционный тест, одновыборочный t-тест.