

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего
образования
«Пермский национальный исследовательский политехнический университет»
Электротехнический факультет
Кафедра: Информационные технологии и автоматизированные системы

Дисциплина: «Математические методы теории систем»
Лабораторная работа № 6
на тему: «Регрессионный анализ в среде R»

Выполнил: студент группы АСУ4-22-1м
Попов Кирилл Михайлович
Проверил: ассистент кафедры ИТАС
Тютюных Артём Александрович

Пермь 2022

ЦЕЛЬ РАБОТЫ

- изучить основные понятия корреляционно-регрессионного анализа;
- выполнить регрессионный анализ в среде R.

ЗАДАНИЕ

1. Выбрать набор данных для анализа и описать значение его переменных;
2. Выполнить регрессионный анализ и получить уравнение модели;
3. Графически отобразить фактические и полученные в результате работы модели данные;
4. Интерпретировать результаты и сделать вывод по проведенному анализу.

ВЫПОЛНЕНИЕ РАБОТЫ

Выбран набор данных: Orange

Фрейм данных имеет 35 строк и 3 столбца записей роста апельсиновых деревьев.

Параметры:

- **tree**: Упорядоченный множитель, указывающий дерево, на котором выполняется измерение.
- **age**: числовой вектор, задающий возраст дерева (в днях с 1968/12/31)
- **circumference**: числовой вектор окружностей ствола дерева (мм). Вероятно, это «окружность на уровне груди», стандартное измерение в лесном хозяйстве.

На рис. 1 приведён код получения регрессионного анализа.

```
> # получение уравнения регрессии
> {
+   lm.tree <- lm(formula = age ~ circumference, data = Orange)
+   lm.tree$coefficients
+   print("уравнение модели:")
+   paste("age = ", lm.tree$coefficients[1], "+", lm.tree$coefficients[2], "*circumference")
+ }
[1] "уравнение модели:"
[1] "age = 16.6036088608175 + 7.81599844386471 *circumference"
> |
```

Рис 1. Регрессионный анализ

На рис. 2 приведён программный код, взятый из примера на датасет, и график к нему (рис. 3). Данный пример описывает изменение возраста дерева по длине его окружности.

```
> # Рост апельсинового дерева
> {
+   require(stats); require(graphics)
+   coplot(circumference ~ age | Tree, data = Orange, show.given = FALSE)
+   fm1 <- nls(circumference ~ sslogis(age, Asym, xmid, scal),
+             data = Orange, subset = Tree == 3)
+   plot(circumference ~ age, data = Orange, subset = Tree == 3,
+        xlab = "Tree age (days since 1968/12/31)",
+        ylab = "Tree circumference (mm)", las = 1,
+        main = "Orange tree data and fitted model (Tree 3 only)")
+   age <- seq(0, 1600, length.out = 101)
+   lines(age, predict(fm1, list(age = age)))
+ }
> |
```

Рис 2. Пример

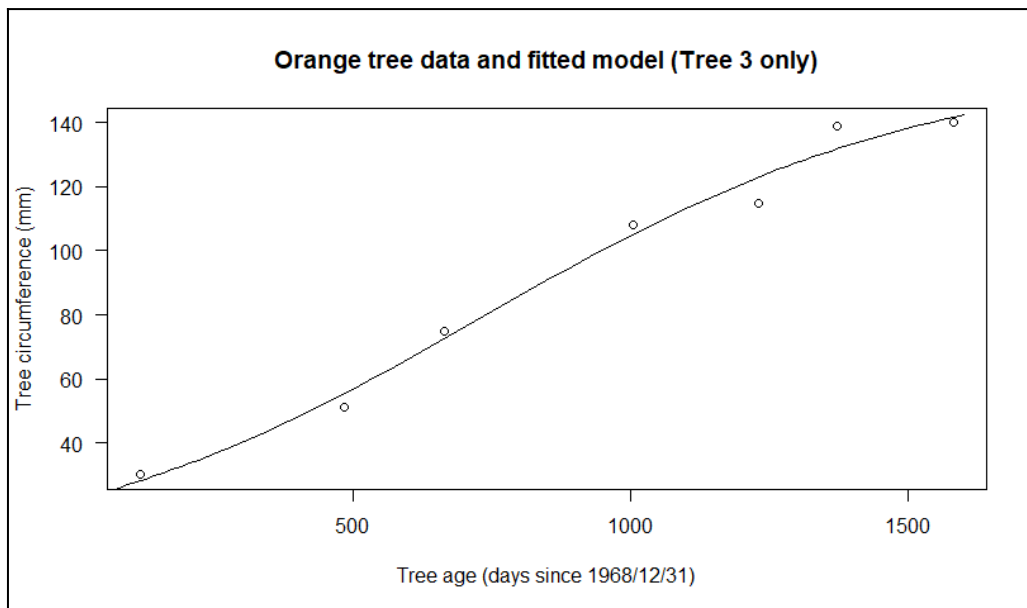


Рис 3. Фактические данные

На рис. 4 и 5 приведены программный код регрессии, описывающий отношение длины окружности ствола апельсинового дерева к времени его роста, и график к нему соответственно.

```
> # Регрессионный анализ набора данных orange
> {
+   b <- lm.tree$coefficients[1]
+   a <- lm.tree$coefficients[2]
+   xmin <- min(orange$circumference)
+   xmax <- max(orange$circumference)
+   x <- seq(from = xmin, to = xmax, length.out = 100)
+   y <- b + a*x
+   plot(
+     orange$circumference,
+     orange$age,
+     main = "Регрессия",
+     xlab = "длина окружности (мм)",
+     ylab = "возраст (дни с 1968/12/31)"
+   )
+   grid()
+   lines(x, y, col="red")
+ }
> |
```

Рис 4. Программный код

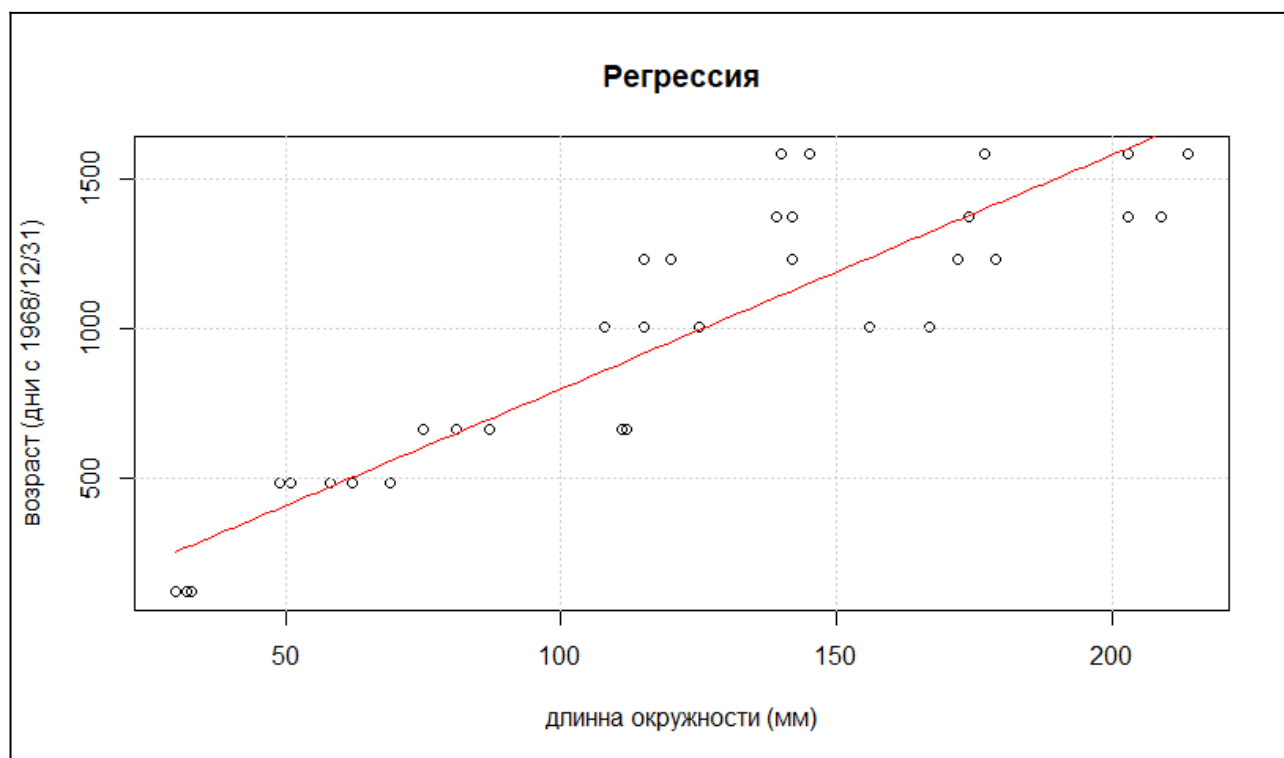


Рис 5. График регрессии

На рис. 6 приведены сведения о составленной модели.

```
> summary(lm.tree)

Call:
lm(formula = age ~ circumference, data = orange)

Residuals:
    Min       1Q   Median       3Q      Max
-317.88 -140.90  -17.20   96.54  471.16

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.6036    78.1406   0.212   0.833
circumference  7.8160     0.6059  12.900 1.93e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 203.1 on 33 degrees of freedom
Multiple R-squared:  0.8345,    Adjusted R-squared:  0.8295
F-statistic: 166.4 on 1 and 33 DF,  p-value: 1.931e-14

> |
```

Рис 6. Сведения о модели

Анализ:

- F-критерий составил 166, что больше 1
- Вероятность получить такое значение при отсутствии связи между x и y (p-value = 1.931e-14).
- Коэффициент детерминации составляет 0.8345

Вывод: большая стандартная ошибка, большой разброс значений. Однако, графики регрессии и фактический данных схожи.

Возможные проблемы: ошибки измерения, недостаток данных, внешние факторы.

ЗАКЛЮЧЕНИЕ

Программный код и отчёт, разработанные в среде RStudio, находятся по адресу:
https://github.com/Kirpo97/MMTS_labs/tree/main/lab_6.