Регрессионный анализ, часть 2. Домашнее задание, вариант 5.

2023-12-02

# Введение

В этой работе мы проверим, можно ли определить возраст моллюсков на основании данных морфометрии, иными словами, существует ли линейная зависимость числа колец на спиле раковины от этих факторов. Это поможет упростить метод подсчёта возраста моллюсков, потому что точный подсчёт колец на спиле их раковины - очень долгий и трудоёмкий процесс [1, 2].

# Методы

В исследовании Nash et al.(1994) [2] были изучены такие морфометрические признаки моллюсков *Haliotis rubra* (на 1310 ЭКЗЕМПЛЯРОВ ), как длина раковины (length), её диаметр (diameter), вес (shell weight), число колец (rings), а также общий вес моллюска (whole weight) и другие в зависимости от их гендера (sex).



Рисунок 1. Внешний вид *Haliotis rubra*

Подбор линейной модели зависимости числа колец на спиле раковины от различных факторов был осуществлён с помощью метода наименьших квадратов [3]. Для тестирования статистической значимости зависимости между исследованными признаками использовался дисперсионный анализ [3], а для оценки условий применимости модели - построение графика зависимости стандартизованных остатков от предсказанных значений [4]. Корректировка модели проведена при помощи преобразования её в модель, основанную на распределении Пуассона (так как зависимая переменная - это счётная величина) [5].

Расчеты проведены в R [6, 7] с использованием функций lm() и glm() из пакета stats [7]; ggplot() и fortify() из пакета ggplot2 [8]; Anova(), residualPlot() и qqPlot() из пакета car [9]; groupp\_by(), do(), ungroup() из пакета dplyr [10] и glm.nb() из пакета MASS [11].

# Результаты и обсуждение

Первая подобранная модель оказалась невалидной из-за гетероскедастичности, поэтому была подобрана модель, основанная на распределении Пуассона (так как зависимая переменная - это счётная величина). Оказалось, что возраст моллюсков (число колец на раковине) *Haliotis rubra* зависит от гендера, диаметра и веса раковины, а также от общего веса моллюска (см. табл. 1).

Таблица 1. Параметры подобранной модели

| Член модели | Параметр | SE | t | p-value |
| --- | --- | --- | --- | --- |
| (Intercept) | 3.495480 | 0.3980015 | 8.7825805 | 0.0000000 |
| sexI | -1.009030 | 0.2053464 | -4.9137958 | 0.0000010 |
| sexM | 0.126143 | 0.1692664 | 0.7452337 | 0.4562646 |
| diameter | 16.361712 | 0.8271126 | 19.7817221 | 0.0000000 |

Эта зависимость описывается уравнением

(см. диаг. 1).

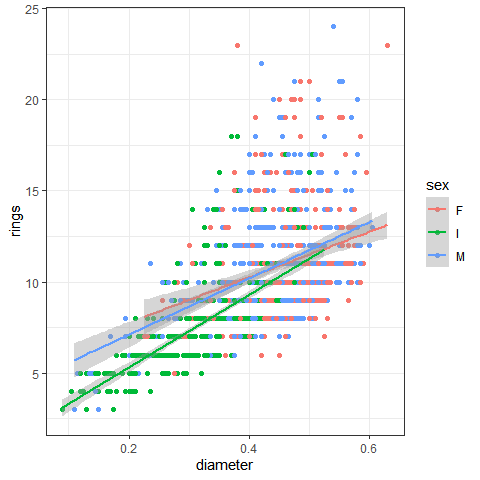


Рисунок 2. Визуализация модели