СОДЕРЖАНИЕ

[Подготовка данных для анализа 4](#_Toc123829094)

[Корреляционный анализ данных 5](#_Toc123829095)

[Регрессионный анализ данных 8](#_Toc123829100)

[коэффициент корреляции пирсона 9](#_Toc123829101)

[Анализ модели множественной регрессии 11](#_Toc123829102)

подготовка данных для анализа

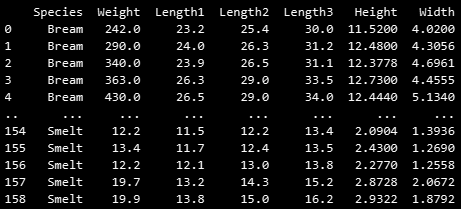


Рисунок 1. Исходные данные для анализа

корреляционный анализ данных

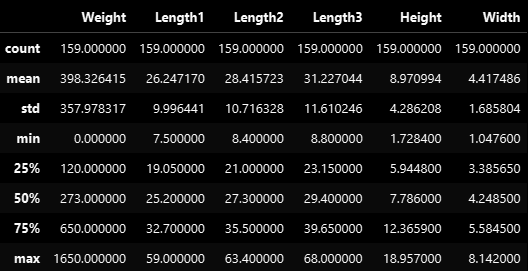


Рисунок 2. Основные показатели описательной статистики

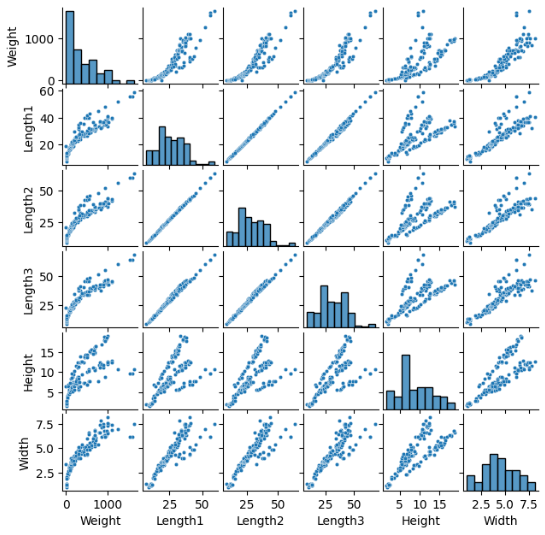


Рисунок 3. Диаграммы рассеяния для выбранных признаков



Рисунок 4. Корреляционная матрица

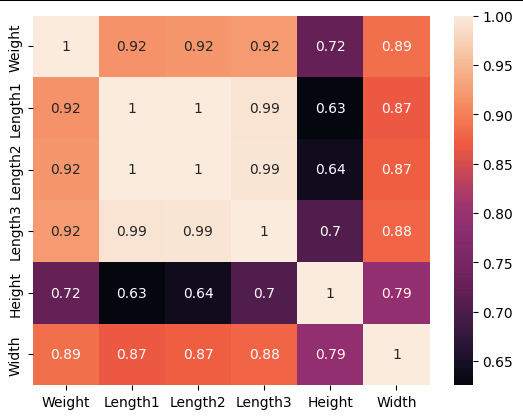


Рисунок 5. «Тепловая карта»



Рисунок 6. Проверка гипотезы о значимости

коэффициентов корреляции

Регрессионный анализ данных

Для пары количественных признаков «Height» – «Length3» коэффициенты уравнения линейной регрессии будут: 0.2597 x + 0.8619 .

residuals (остатки модели) = 9.22323645109217 .

Глядя на графическое представление модели, можно заметить, что не так хорошо подходит для описания имеющихся данных. При анализе графика зависимости остатков модели наблюдается U-образная фигура, что указывает на нелинейный характер зависимости y от х.

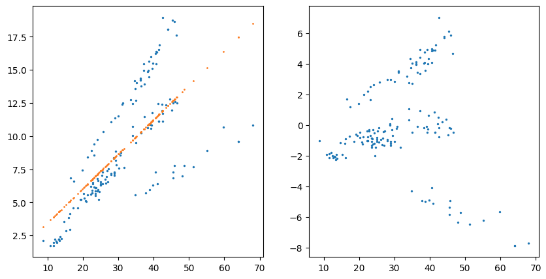


Рисунок 7. Диаграмма остатков линейной

регрессии и дисперсия остатков

Для пары количественных признаков «Height» – «Length3» коэффициенты уравнения полиномиальной регрессии будут: 5.807e-07 x5 - 9.318e-05 x4 + 0.005217 x3 - 0.128 x2 + 1.758 x - 7.535

По мере увеличения степени полинома дисперсия увеличивается незначительно. Также видим положительную взаимосвязь.

residuals = 6.027084159312364

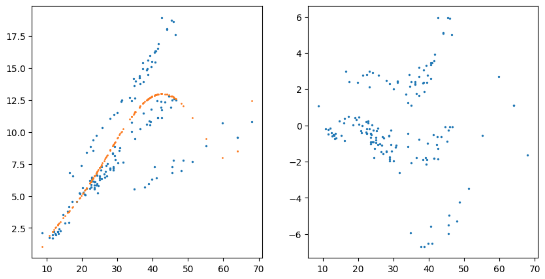


Рисунок 8. Диаграмма остатков полиномиальной

регрессии и дисперсия остатков

коэффициент корреляции пирсона

Индекс корреляции больше, чем 0.5, что означает высокую силу связи между «Length3» и «Height», что совпадает с коэффициентом корреляции, расcмотренный ранее.

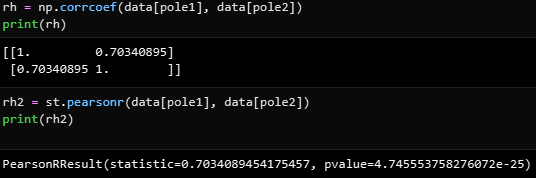


Рисунок 9. Индекс корреляции

Анализ модели множественной регрессии

Отклик – Length3, переменные факторы – Width, Height.

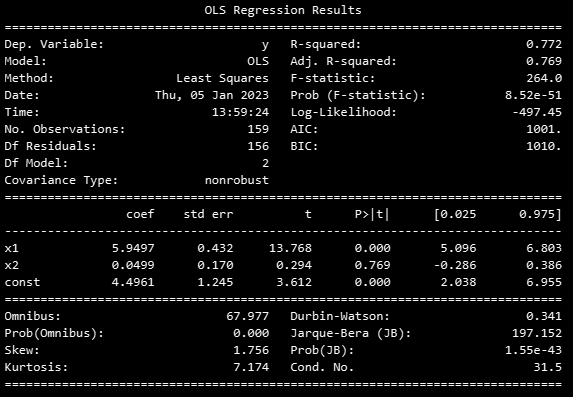


Рисунок 10. Модель множественной регрессии

Откликом является Length3 рыбки. Модель: OLS (минимизация суммы квадратов). 159 наблюдений всего. 156 независимых наблюдений.

const → 4.4961 - среднее ожидаемое значение высоты при нулевых факторах.

coef → показывает скорость изменения отклика y при изменении фактора x1 (5.9497), x2 (0.0499). Получается, что Length3 меняется быстрее при изменении ширины, чем при изменении высоты.

st err - это стандартная девиация.

P>|t| → высота (0.769) больше 0.05, т.е. статистически незначима связь между высотой и длиной (Length3) рыбы.

R-squared – 0.772, модель довольно хороша. Чем ближе к 1 тем лучше.

Adj. R-squared – 0.769. Сильно не отличается от R-squared, значит, что выбранные признаки являются актуальными и соответствуют модели.

F-statistic и p-value – Поскольку F-statistic (>1), намного больше p-value а p-value близка к нулю (<0.05), то это говорит о линейной зависимости между признаками и откликом.