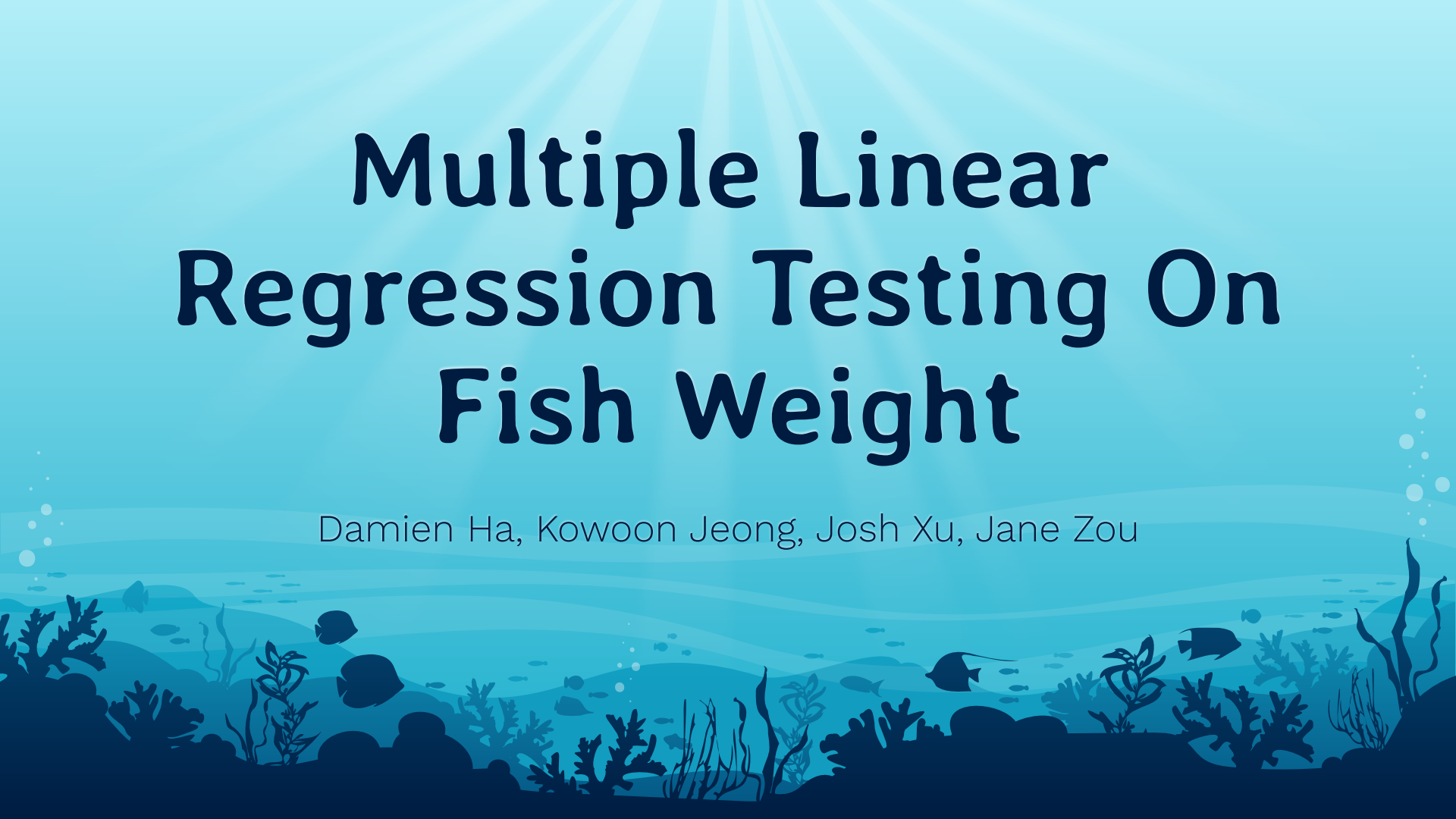
Multiple Linear Regression Testing On Fish Weight

Damien Ha, Kowoon Jeong, Josh Xu, Jane Zou



Background / Data Source

- **Research Question:** predictive MLR model to study how length measurements affect fish weight
- Kaggle, 7 species in fish market sales, 159 x 7
- Fish weight explains fishery health and proper management

Variables

- Species: (categorical) bream, parkki, perch, pike, roach, smelt, whitefish

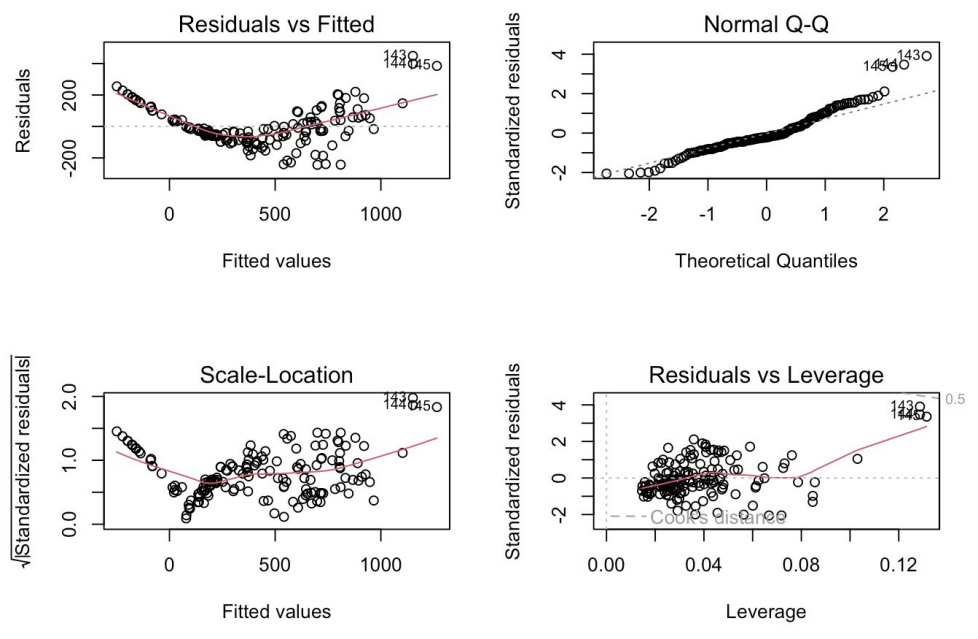
Response

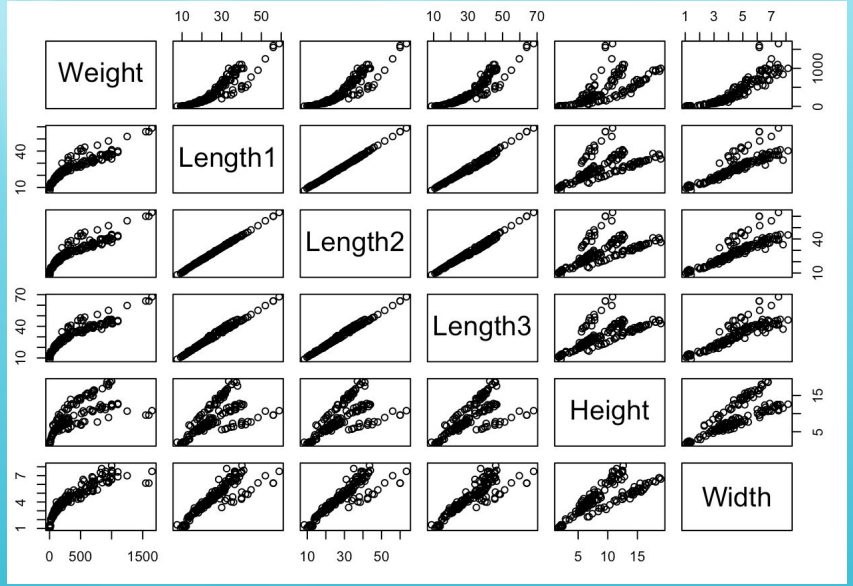
- Weight: (g)
  - Removed 0-weight fish in row 41

Explanatory

- Length1: Vertical (cm)
- Length2: Diagonal (cm)
- Length3: Cross (cm)
- Height: (cm)
- Width: Diagonal (cm)

Linear Regression





Length1	Length2	Length3	Height	Width
1677.71215	2082.43649	421.83683	14.55657	12.26187

$$\text{Weight(hat)} = -496.802 + 63.969\text{Length1} - 9.109\text{Length2} - 28.119\text{Length3} + 27.926\text{Height} + 23.412\text{Width}$$

Power Transformation, Variable Selection

- Multiple R^2 = 0.9947; Adjusted R^2 = 0.9945
  - All possible subsets remove Length1
  - Partial F-test for stepwise remove Length3

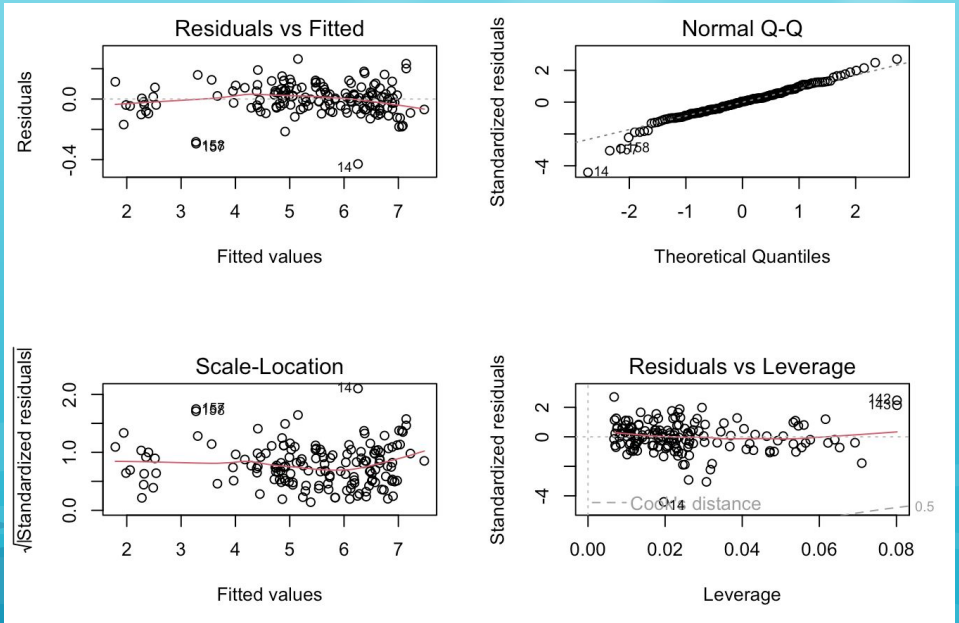
Models with these variables were selected for each method; note that logs have been taken for each variable:

- Forward stepwise AIC: Length2, Length3, Height, Width
- Forward stepwise BIC: Length2, Length3, Height, Width
- Backward stepwise AIC: Length2, Height, Width
- Backward stepwise BIC: Length2, Height, Width

$$\log(\text{Weight}) = -1.942 + 0.401\log(\text{Length1}) + 1.596\log(\text{Length2}) - 0.513\log(\text{Length3}) + 0.68\log(\text{Height}) + 0.845\log(\text{Width})$$

Final Model

$$\log(\text{Weight}) = -2.01 + 1.498\log(\text{Length2}) + 0.612\log(\text{Height}) + 0.902\log(\text{Width})$$



tLength2	tHeight	tWidth
7.666024	5.507957	14.023230

Multiple $R^2 = 0.9946$; Adjusted $R^2 = 0.9945$

Conclusion

- Studies support log transformation to predict fish weight based on length measurements
- **Limitations**: Lack data on certain fish species, ignored Species column
- **Future Improvements:** fit individual models for each species with different predictors