# Model for Predicting Body Fat Percentage (Brozek's Equation)

## Overview

The goal was to develop a model for predicting Body Fat Percentage using only simple physical measurements like weight, height, circumferences of certain body parts, and age. These variables were chosen because they do not require additional calculations (e.g., Adiposity Index = Weight/Height²). Brozek's equation was used to calculate the Body Fat Percentage.

## Analysis

After data cleaning and preparation, I wanted to find out how predictor variables are related to the response variable. Several predictor variables showed a linear relationship with Brozek's Body Fat Percentage, with Density having a near-perfect (-0.99) correlation. However, variables like Adiposity Index and Fat-Free Weight, which required additional calculations, were excluded as I wanted to build a model that uses only simple measurements that can be taken using only a tape and a scale. A linear regression model using only Density as the predictor yielded an adjusted R2R^2R2 of 0.9993, indicating excellent predictive power. Still, the focus was to build a model with physical measurements.

A linear model using all 13 variables, excluding Density, Adiposity Index, and Fat-Free Weight, gave an adjusted R^2 of 0.7302. Subset regression further reduced the model to 8 predictors, improving the adjusted R^2 to 0.7337. Testing this model on a split dataset resulted in a test R^2 of 0.7099,

Given the high multicollinearity among variables, Ridge and Lasso regression were employed, as they are better suited for handling multicollinearity. Both models were trained using the original 13 variables. Ridge regression achieved a test R^2 of 0.6990, while Lasso regression slightly outperformed with a test R^2 of 0.7181.

## Conclusion

Lasso regression emerged as the most effective model for predicting Body Fat Percentage based on its slightly higher test R^2 value. However, the linear regression model also performed comparably, making it a viable option and both models do not use all the predictor variables which makes them simpler. It seems all three models give fairly solid results for predictions considering the fact I did not include the predictor variable Density which has the highest correlation with the body fat percentages and other variables that are highly influential like Fat_Free_Weight.

Coefficients for Lasso Regression;

| | |
|---|---|
| (Intercept) | 4.47612085 |
| Age | 0.06588172 |
| Weight | . |
| Height | -0.29301184 |
| Neck | -0.51444378 |
| Chest | . |
| Hip | . |
| Abdomen | 0.66640151 |
| Thigh | . |
| Knee | . |
| Ankle | 0.01525737 |
| Biceps | . |
| Forearm | 0.30347815 |
| Wrist | -1.04748018 |

Coefficients for linear Regression;

| (Intercept) | Age | Weight | Neck |
|---|---|---|---|
| -21.71465449 | 0.07594145 | -0.09730913 | -0.71801307 |
| **Hip** | **Abdomen** | **Thigh** | **Forearm** |
| -0.16444999 | 0.90961947 | 0.28990486 | 0.57313386 |
| **Wrist** | | | |
| -1.07069336 | | | |