Appendicular Lean Mass: Insights through Regression Analysis

Prof. Peter Wolenski and Dr. Nadejda Drenska PALM Team- Amaya Evans, Anna Crifasi, Artem Mukhamedzianov, Atif Iqbal, Gowri Sunkara, Gyaneshwar Agrahari, Iswarya Sitiraju, Serene Sam, Teddie Swize

Department of Mathematics, Louisiana State University

Pennington Biomedical Research Center
Louisiana State University

Introduction

The LSU Department of Mathematics is working jointly with Pennington Biomedical Research Center (PBRC) to find alternatives to dual energy x-ray absorptiometry (DXA) scans. DXA scans are a costly method of measuring body composition variables such as Appendicular Lean Mass (ALM), body fat percentage, and bone density. These components are important for research in osteoporosis, obesity, and healthy aging (CDC 2021).

Objective

The objective of this research project is to apply linear regression, polynomial regression, and principal component analysis (PCA) to predict appendicular lean mass (ALM). The primary aim is to conduct a comparative analysis of these methodologies, evaluating their predictive performance using mean absolute, root mean square, and maximum error metrics. The models involve a diverse set of anthropometric measurements and demographic variables, distinguished by sex.

Methodology

Pennington provided a dataset encompassing information on 846 patients, featuring age, which ranged from five to eighty-nine years old, sex, race, and forty-four biomarkers along with DXA-measured ALM values.

Data Preprocessing: Three datasets were created based on sex—Male, Female, and Combined. Patients with missing measurements were excluded. The average ALM measured by DXA was 22.8 kg for males, 16.1 kg for females, and 19.3 kg for the combined set.

The datasets were then subdivided and tested with and without the use of Principal Component Analysis (PCA). PCA was utilized to reduce dimensionality and identify influential biomarkers.

Regression Methods:

1. Linear Regression (LR): A statistical approach used to model the relationship between the ALM and the biomarkers by fitting a linear equation to the observed data
2. Polynomial Regression (PR): Similar to linear regression, but a degree two polynomial is used to fit the observed data.

Testing: To evaluate model accuracy, a representative subset consisting of 50 randomly selected males and 50 females was employed. The combined set was tested on the union of the two previously mentioned sets.

Evaluation Metrics: The accuracy of each model was rigorously assessed using three distinct error metrics:

$$\text{Mean Absolute Error: } \frac{1}{N}\sum_{k=1}^{N}|\hat{y}_k - y_k|$$

$$\text{Root Mean Squared Error: } \sqrt{\frac{1}{N}\sum_{k=1}^{N}|\hat{y}_k - y_k|}$$

$$\text{Maximum Error: } \max_{k=1,2,\ldots,N}|\hat{y}_k - y_k|$$

y_k: True ALM of patient x_k
$\hat{y}_k$: Predicted ALM of patient x_k

Results

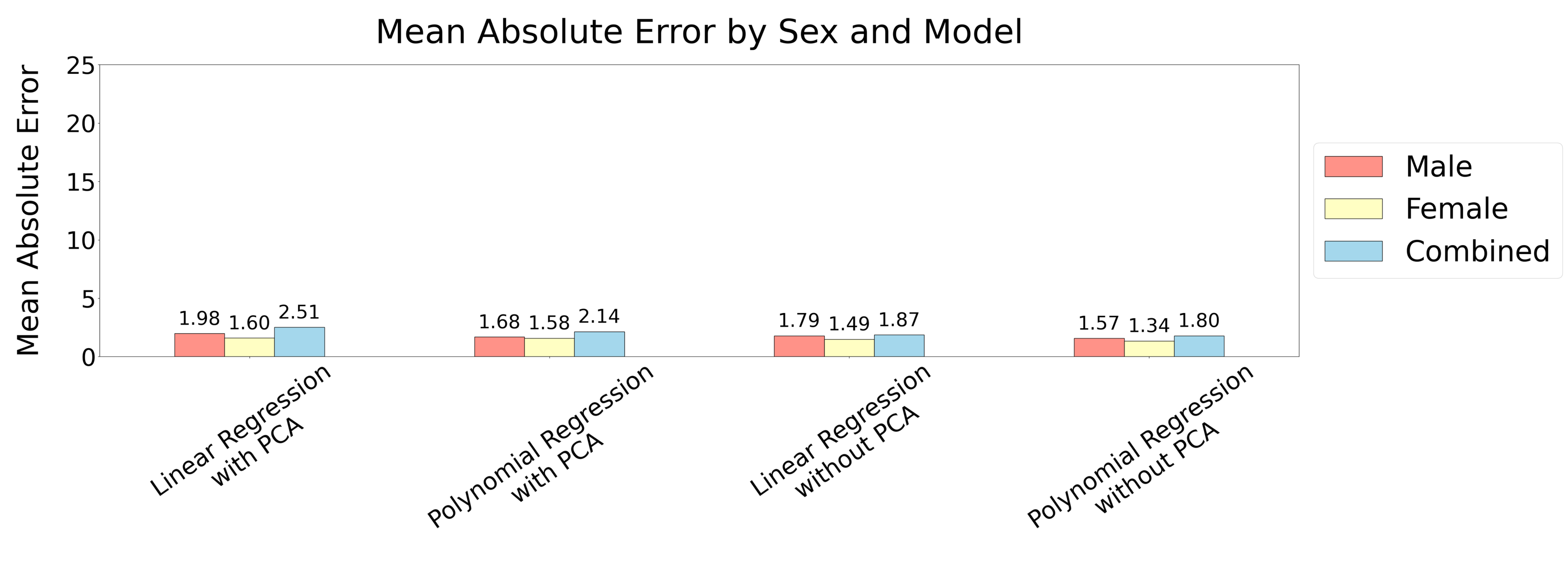

Figure 1.

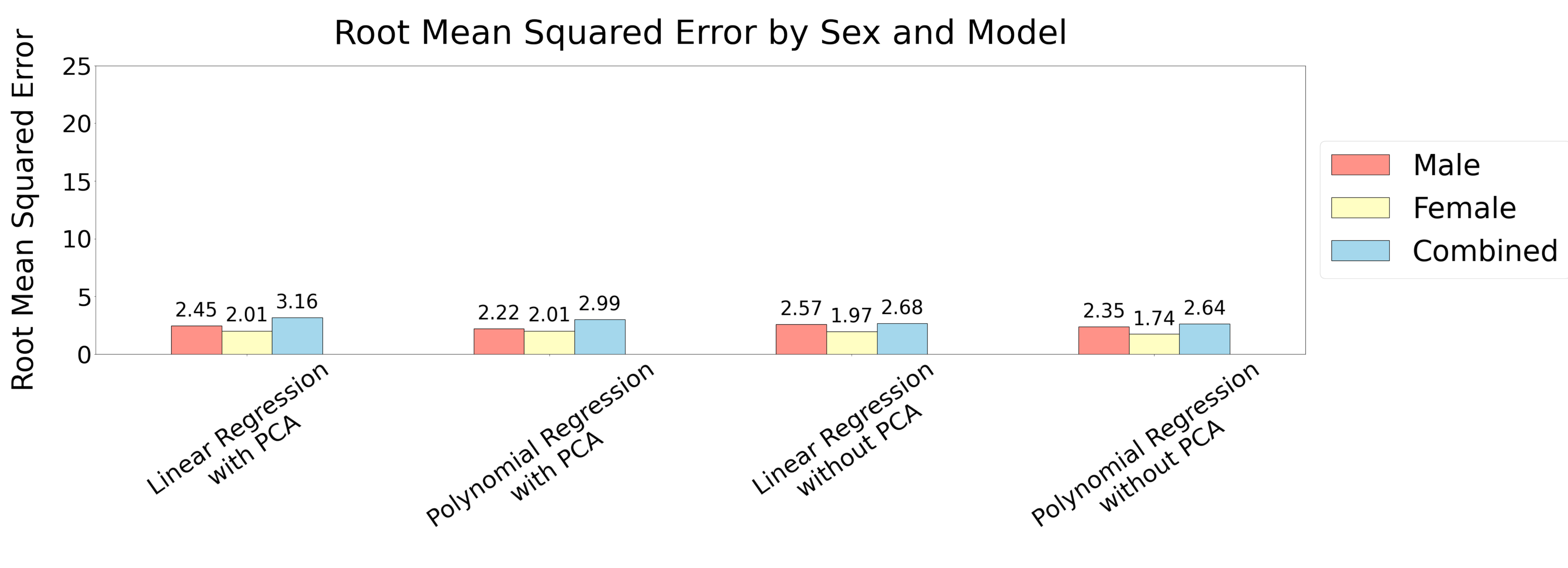

Figure 2.

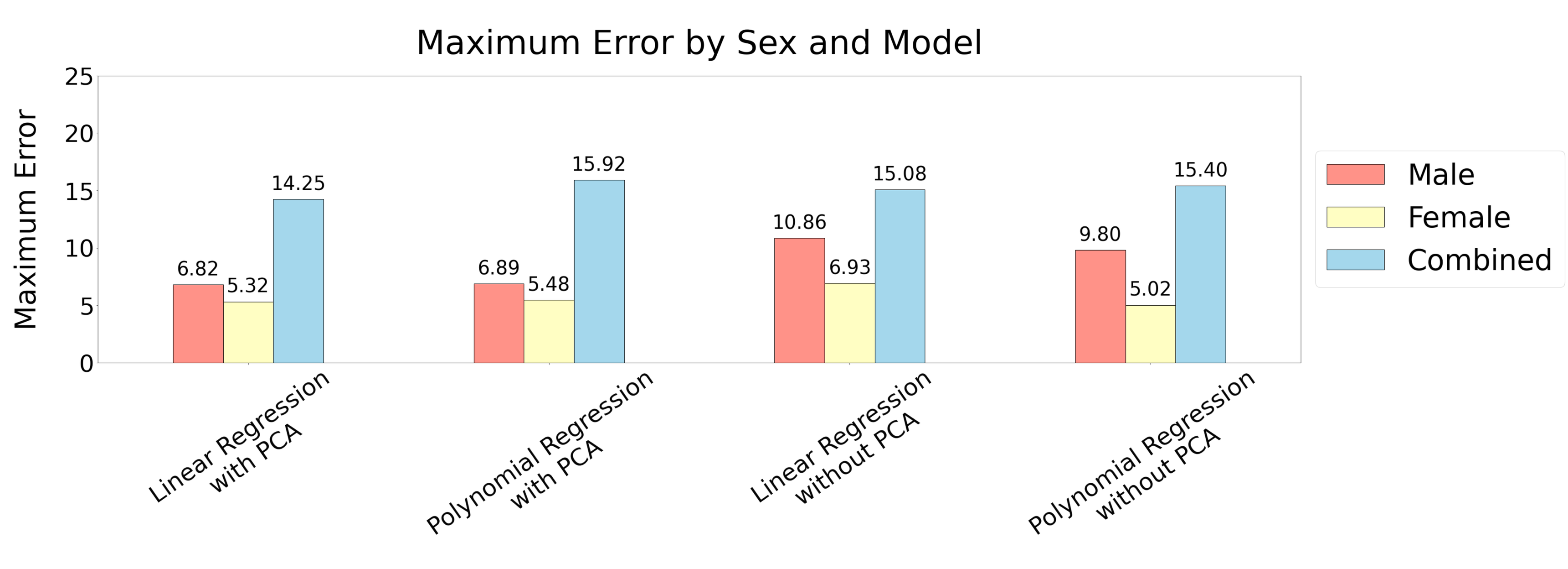

Figure 3.

Discussion

The graphs demonstrate that first dividing the dataset into male and female subsets and then applying the models improved the efficiency. The maximum error of the combined set was significantly greater than both the male and female subsets. This finding suggests that sex plays an important role in providing more accurate ALM predictions.

On the other hand, the use of PCA did not greatly improve efficiency. As seen in the table below, the models with PCA overall performed worse than those without.

Dataset	LR with PCA vs LR	PR with PCA vs PR
Male	5%	6%
Female	-2%	-15.5%
Combined	-17%	-13%

Table 1. PCA Performance: Percent Comparison

Although PCA did not increase model accuracy, it did prove to be beneficial in identifying influential variables. The following are the principal components of each subset:

- **Male**—head circumference; right forearm circumference
- **Female**—head circumference; left thigh circumference
- **Combined**—head circumference; right leg volume

As a principal component of all three subsets, head circumference was found to be the most important factor in predicting ALM.

Best Fit Model: As expected, the polynomial regression achieved a smaller root means square error than the linear regression, meaning PR renders a more accurate model.

Future Work

Future studies may build upon these models to predict various DXA measurements, such as bone density and body fat percentage. Additionally, we plan to apply a semi-supervised learning algorithm based on the p-Laplacian. We hope that our models will be integrated in future obesity, metabolism, and nutrition studies.

Acknowledgements

- We would like to thank Professor Peter Wolenski and Dr. Nadejda Drenska from the Department of Mathematics, Louisiana State University, for putting together the whole team and helping make this project a success.
- Department of Mathematics, Louisiana State University.
- Pennington Biomedical Research Center, LSU.

References

- Center for Disease Control and Prevention. (2021, October 20). *Radiation in Healthcare: Bone Density (DEXA Scan).* https://www.cdc.gov/nceh/radiation/dexa-scan.html