Executive Summary

Introduction:

The utilization of DEXA and other X-ray machines to measure body fat percentage require a great amount of time and money. Therefore, our goal in this project is to create a more affordable and accessible, but still fairly accurate, estimation of a person's body fat percentage through linear regression.

Data Cleaning:

We noticed that individual 182 had a body fat of 0% which is highly unlikely because even athletes do not have a body fat percent that low so we used Siri's equation to recalculate this individual's body fat percent. However plugging this individual's density into Siri's equation, resulted in a negative value for body fat percent so we deleted this observation. We also noticed that individual 42's height is only 29.5 inches which is quite abnormal. We used the BMI formula to recalculate the individual's height. We concluded that it was a recording error and fixed his height to 69.43 inches.

Our final model:

$$\text{Body Fat Percentage} = -23.934 + 0.898\text{Abdomen} - 0.106\text{Weight} - 1.170\text{Wrist}$$

Example Usage: A man with an abdomen circumference of 95.6 cm, weight of 177 pounds, and wrist circumference of 17.7 cm is expected to have a body fat percentage of 22.52 .

Rule of thumb: "multiply your abdomen circumference by 0.9 , multiple your weight by -0.1, multiple you wrist circumference by -1.2, add all three terms together, then subtract 24."

Example Usage: Using the rule of thumb , a man with an abdomen circumference of 95.6 cm, weight of 177 pounds, and wrist circumference of 17.7 cm is expected to have a body fat percentage of 24.87.This rule of thumb overestimates the percentage of body fat compared to the model with the non-rounded coefficients.

Model Interpretation:

- The average body fat percentage decreases by 0.106 for a one pound increase in weight.
- The average body fat percentage decreases by 1.146 for every additional cm in wrist circumference.
- The average body fat percentage increases by 0.902 for every additional cm in abdomen circumference.

Model Choice: According to an article from Healthline, the more visible your abdominal muscles are the lower your body fat percentage so we decided to start off with a simple linear regression(SLR) model with just abdomen as a predictor. We found that abdomen was statistically significant at $\alpha = 0.05$ with a p-value of $<2\times10^{-16}$. We added different combinations of predictors to the SLR model and chose the model with significant predictors and highest R-squared. We found that weight is statistically significant at $\alpha = 0.05$ and its model also had the highest R-squared among the other combinations. Finally, we decided to try adding a third predictor to our model by also trying different combinations of predictors. We added wrist as a third predictor because it is statistically significant at $\alpha = 0.05$ and its model had the highest R-squared. We had three candidate models and noticed that the R-squared value increased from the SLR model to our final model. Our final model with abdomen, weight and wrist as predictors explains about 72.04% of all the variation in body fat percentage which was the highest of the three models. Our final model also has a low RMSE value of 6.747 indicating a better fit than the other two models.

Model Diagnostics: Before using our model for prediction, it is crucial to verify that our final model does not violate the assumptions of the linear model. The assumptions we diagnose include (1) Linearity ,(2) Homoscedasticity, (3) Normality, and (4) Independence.

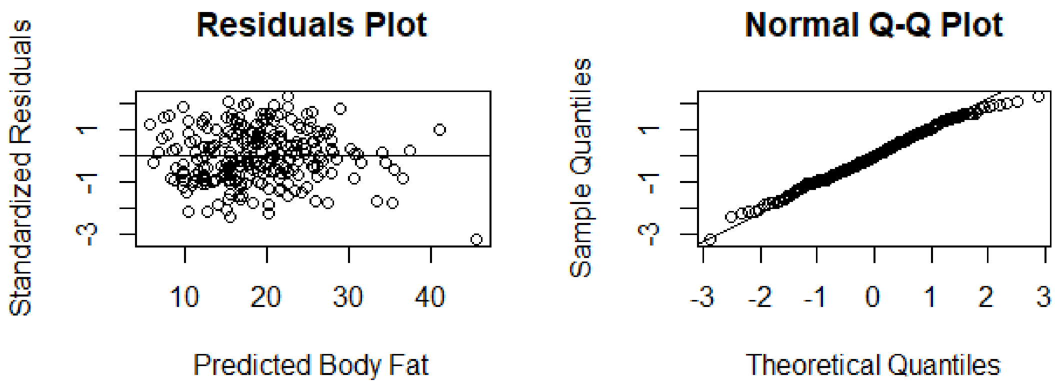

Figure 1: Residuals Plot and Normal Q-Q plot

Overall, all assumptions of the linear model are satisfied and our proposed model is appropriate. Even though there are a few observations that have relatively high residuals,the residual plot (see Figure 1) shows no distinctive pattern and is evenly spread around $y = 0$ line. Therefore, the linearity and homoscedasticity as well as the independence assumptions hold. It is also plausible to confirm that the normality assumption is satisfied because most of the observations in the Q-Q plot lie along the 45-degree line (also see Figure 1). The p-value from the Shapiro-Wilk test also confirms our conclusion that the normality assumption at $\alpha = 0.05$ holds due to a large p-value

Model Strengths and Weaknesses: Some strengths of our model include the fact that it is intuitive, easy to build, and significantly accurate. Our model also satisfies all the four assumptions of linear regression and explains about 72.04% of all the variation in body fat, this brings credence to our model and final results not being too overfitting. Some weakness of our model is that we include an observation with a weight of 363 so we sacrifice having a higher R-squared value and accuracy. However, we do get a better understanding of the data and prevent overfitting. Our model also accurately predicts body fat percentage of those that fall in the data range. Also, it is less reliable when predicting a female's body fat.

Conclusion:
In summary, we fit a multiple linear regression model with three selected predictors: Abdomen, Weight and Wrist. The model satisfies the linear model assumptions and accurately predicts the body fat.Even though there are some limitations, the model is user-friendly and accurate for the data range.

Contributions:
ME: Worked on the Model/Model Choice in both the executive summary and powerpoint , and edited the final executive summary.
SH: Created the ShinyApp and worked on the Introduction/Data cleaning in both the executive summary and powerpoint.
SS: Created the GitHub repository and worked on the Model Diagnostics in both the executive summary and powerpoint, and edited the powerpoint
XW: Worked on the Model Strengths & Weakness and Conclusion in both the executive summary and powerpoint and uploaded all our files to GitHub.

Together: We cleaned the dataset, coded our potential models, and edited the executive summary together.

References:

Lindberg, S. (2018, September 27). *What's the ideal body fat percentage for ABS?* Healthline. Retrieved October 7, 2022, from https://www.healthline.com/health/body-fat-percentage-for-abs

Home. Calculator.net: Free Online Calculators - Math, Fitness, Finance, Science. (n.d.). Retrieved October 7, 2022, from https://www.calculator.net/body-fat-calculator.html