Stat 628: Module 2 - Group 8 Executive Summary

Introduction

It is desirable to have an accurate method of measuring body fat. Using a dataset with 252 men's body measures, we built a simple, robust, "rule of thumb" to estimate body fat percentage. Given a person's weight and abdomen measure we can accurately estimate body fat percentage.

Data Cleaning

Initial data analysis showed points as erroneous. A recorded body fat percentage of 0% was attempted to be recovered utilizing the density. When re-calculated with the formula specified by Siri [1] the body fat was -3.6%, being impossible this row was removed. A point had a height of 29 inches, which is infeasible. The height was recovered using the recorded BMI and weight and was found to be 69.43 inches. An outlier abdomen measurement was removed for the model.

Model

Our proposed model for estimating body fat %:

$$Body\ Fat\ \% \ = \ -42.50 \ + \ 2.293 * Abdomen(in) \ - \ 0.123 * Weight(lbs)$$

Our proposed rule of thumb for estimating body fat %: "Double your abdomen in inches, subtract 10% of your weight in pounds, and subtract 40."

For example, an individual with an abdominal girth of 33 inches and 150 pounds will have an estimated body fat percentage of 14.72%. The coefficients can be interpreted by seeing how body fat % changes while holding the other variable constant. For abdomen circumference we expect an increase of about 2.3% body fat for a one-inch increase in circumference while holding weight constant. Similarly for weight for every 10 pounds gained while holding the abdomen constant an individual will lose 1.23% body fat.

Statistical Analysis

For model selection, we evaluated the adjusted R-squared of linear models with three or fewer variables. We used the adjusted R-squared due to modeling with different numbers of variables. We wanted a linear model because it is easy to interpret and leads to an easy-to-use rule of thumb. We investigated other body fat % calculators and formulas to narrow our potential variables. We decided to try models with a combination of adiposity, height, weight, and abdominal circumference. Figure 3 shows a bar chart of the top 5 models by adjusted R-squared when considering only the previously specified variables.

Although the best model included abdomen, height, and weight, this only changed our adjusted R squared slightly (0.7184 to 0.717) with significant statistics of weight and abdomen. In the interest of creating a simple rule of thumb, we decided to use only weight and abdomen so the audience would need only 2 measures.

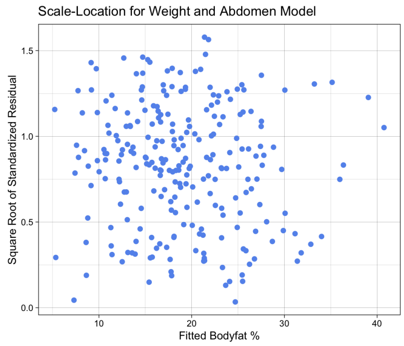

Figure 1

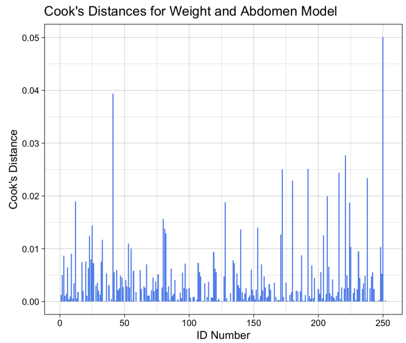

Figure 2

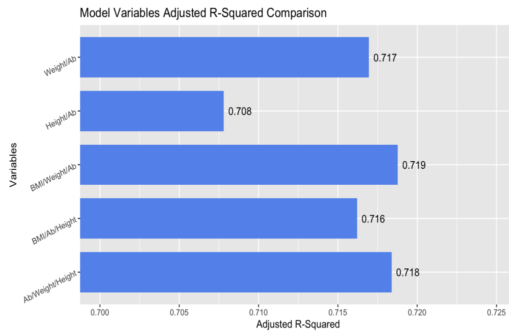

Figure 3

Model Diagnostics and Analysis

In Figure 1, The square root of the residuals appears to be randomly scattered with no strict trend and are in the same range as the estimated values get larger. This indicates our model assumption of homoscedasticity of residuals holds. Figure 2 shows an estimate of the influence of a data point to the estimated coefficients, the Cook's distances of the data points are in the reasonable range. Originally one data point was much larger with a high amount of leverage, and as noted, was removed for the final model. Given these diagnostic plots our simple linear model appears to be appropriate for our rule-of-thumb to calculate body fat percentage.

The model r-squared shows that weight and abdomen measure explain 71.7% of the variation in body fat percentage. We are 95% confident that the true slope for weight is in (-0.162, -0.085) and abdomen is in (2.034, 2.552). The coefficients for both weight and abdomen were found to be significant with p-values of $1.39\text{x}10^{-9}$ and less than $2\text{x}10^{-16}$ respectively.

Strengths: Our model is simple, and its parameters can be easily interpreted. It has a high adjusted-r squared of 0.717. The summary plots indicate that the assumptions of the model, such as homoscedasticity, are met so our model is reasonable.

Weakness: Our model is a simple linear model with two variables. We believe that increasing the complexity of the model might further increase the model's ability to generalize. The model lacks the ability of applicability to the wider population. The data used in this study came from males ranging from ages 22 to 81. Additionally, all the samples used in the final model are from individuals weighing between 115 and 265 pounds. Therefore, if the model is used for individuals outside of the age or weight range it is likely our model becomes far less accurate.

Conclusion

Our model can be used to estimate body fat percentage from a person's weight in pounds and abdomen in inches. Our model diagnostics show plausible linearity and homoscedasticity of the residuals. However, our model is sensitive to individuals outside of the specified weight range.

Works Cited

Centers for Disease Control and Prevention. (2022, September 2). Adult BMI calculator. Centers for Disease Control and Prevention. Retrieved October 17, 2022, from https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/english_bmi_calculator/bmi_calculator.html

[1] Siri, William E. "Body composition from fluid spaces and density: analysis of methods." (1956).

Member's Contribution to the Summary, the Presentation, and the Code

AH: Wrote/edited the Introduction, Model Diagnostics Analysis and Conclusion parts of the summary. Worked on slides 2,3,5,8 for the presentation. Created code related to the final data model. Edited summary report and presentation for grammatical errors and clarity.

OH: Wrote Shiny App and maintained GitHub repo. Wrote/edited the Data Cleaning, and Model parts of the summary. Worked on slides 3,4,5,7,9 for the presentation. Created code related to data cleaning and edited R file for final graph creations. Edited summary report and presentation for grammatical errors and clarity.

YL: Wrote/edited The Statistical Analysis, Model Diagnostic and Analysis parts of the summary. Worked on slides 5,6,10 for the presentation. Created code related to data cleaning and variable selection. Edited summary report and presentation for grammatical errors and clarity.