Stat 4110

Data Analysis Summary

Overview:

This analysis explores the intricate connections between various physical measurements and their impact on the estimation of body fat percentages using Brozek and Siri formulas and decides the best model for predicting body fat. The dataset has a range of measurements from individuals, focusing on key variables like height, weight, thigh, bicep, neck, and forearm circumferences. These specific measurements were selected due to their prominence in the final models obtained from stepwise regression analysis.

Analysis:

After thorough data cleaning, the study delved into the dataset with descriptive statistics revealing interesting insights. For instance, the mean body fat percentages were 19% (Brozek) and 19.21% (Siri), indicating a consistency in body fat estimation across methods. The age of participants ranged from 22 to 81 years, with a median of 43 years, suggesting a diverse age representation. The correlation analysis highlighted notable associations between body measurements and body fat percentages. An interesting discovery was the Brozek model had the highest correlation with the thigh, but it was not included in the stepwise regression. Next, Linear regression models further refined these findings. We found the best model was Brozek and included variables like density, weight, height, adiposity, fat-free weight, neck, and forearm measurements. This model achieved an exceptionally high R-squared value of 0.9994 for Brozek, demonstrating their effectiveness in predicting body fat percentage. Categorical analysis divided the dataset into age groups, revealing how body fat distribution varies with age.Between the ages of 18-28 the Brozek body fat was higher than the Siri body fat. Also, above the age of 39, we see a trend where the Siri body fat is calculated to be higher than the Brozek model. The histograms and scatter plots of measurements like thigh and bicep circumferences against body fat percentages provided visual insights into these relationships and their linearity.

Conclusions:

The analysis successfully identifies critical measurements that significantly impact the estimation of body fat percentage. The thigh circumference, in particular, emerged as a vital predictor, but interestingly it was not included in the best Brozek model. We found the best model for predicting body wait was using the Brozek formula with the variables: density, height, weight, adiposity, fat free weight, neck, and forearm. These insights could be instrumental for health professionals in evaluating body fat more accurately and tailoring health and fitness recommendations.

Siri Stepwise Model:

```
> Johnny Ferrara
Error: unexpected symbol in "Johnny Ferrara"
> summary(stepwise_model_Siri)

Call:
lm(formula = BodyFat_Siri ~ Density + Weight + Height + Adiposity +
    FatFreeWeight + Thigh + Biceps, data = my_data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7631 -0.0986 -0.0182  0.0625  3.3427

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.221e+02  6.805e+00  62.035  < 2e-16 ***
Density       -3.916e+02  4.927e+00 -79.471  < 2e-16 ***
Weight         2.550e-02  7.602e-03   3.354 0.000926 ***
Height         1.524e-01  3.893e-02   3.915 0.000118 ***
Adiposity      2.040e-01  5.193e-02   3.929 0.000111 ***
FatFreeWeight -7.027e-02  6.900e-03 -10.185  < 2e-16 ***
Thigh         -1.186e-02  7.559e-03  -1.569 0.118023
Biceps         2.009e-02  1.083e-02   1.856 0.064729 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2919 on 241 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9988
F-statistic: 2.912e+04 on 7 and 241 DF,  p-value: < 2.2e-16
```

Brozek Stepwise Model:

```
>
> Johnny Ferrara
Error: unexpected symbol in "Johnny Ferrara"
> summary(stepwise_model_Brozek)

Call:
lm(formula = BodyFat_Brozek ~ Density + Weight + Height + Adiposity +
    FatFreeWeight + Neck + Forearm, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.66176 -0.07588 -0.02224  0.04199  1.92490

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)    3.880e+02  4.494e+00   86.334  < 2e-16 ***
Density       -3.605e+02  3.285e+00 -109.732  < 2e-16 ***
Weight         2.196e-02  5.196e-03    4.226 3.38e-05 ***
Height         1.571e-01  2.609e-02    6.020 6.45e-09 ***
Adiposity      2.029e-01  3.503e-02    5.793 2.15e-08 ***
FatFreeWeight -6.841e-02  4.609e-03  -14.841  < 2e-16 ***
Neck           1.811e-02  9.823e-03    1.844   0.0664 .
Forearm        1.634e-02  8.466e-03    1.930   0.0547 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1977 on 241 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9994
F-statistic: 5.448e+04 on 7 and 241 DF,  p-value: < 2.2e-16
```