Body Fat: Can we calculate it? Yes we Can!

Peter Kryspin, Catherine Zheng, Jon Starfeldt

Rule of Thumb

For adult males: Body Fat Percentage =

Abdomen Circumference (cm)*.6 + Age*.1 - 40

Confidence Intervals for each Parameter:

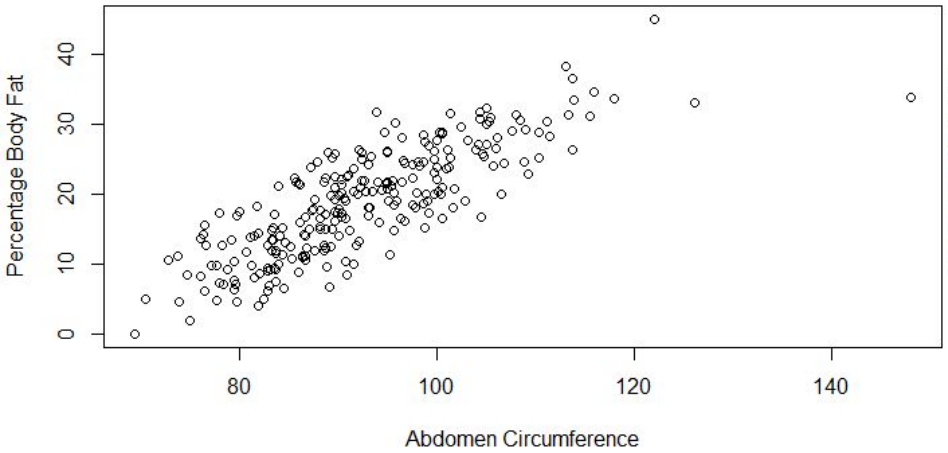
	2.5%	97.5%
Intercept	-44.10	-34.23
Abdomen	0.55	0.65
Age	0.02	0.10

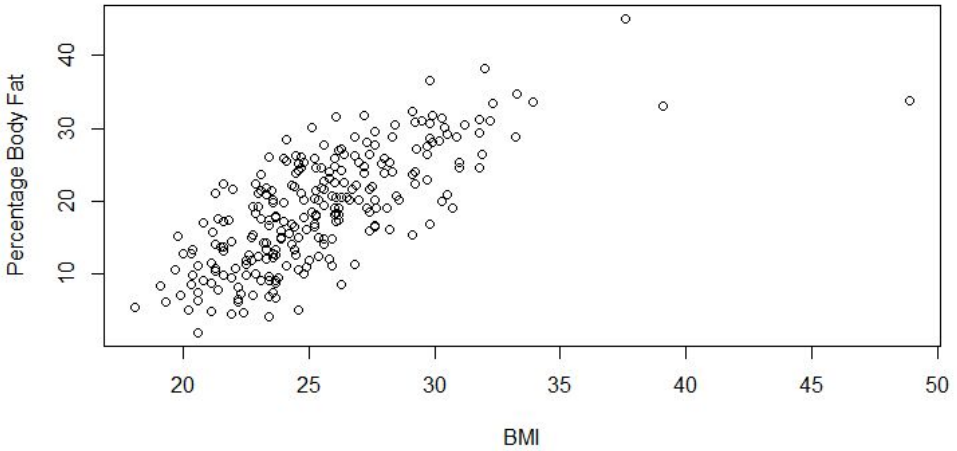
For Example

A man who is 20 years old and has an abdomen circumference of 90.0 cm is predicted to have

90(.6) + 20(.1) - 40 = 16% Body Fat

Initial Model Creation



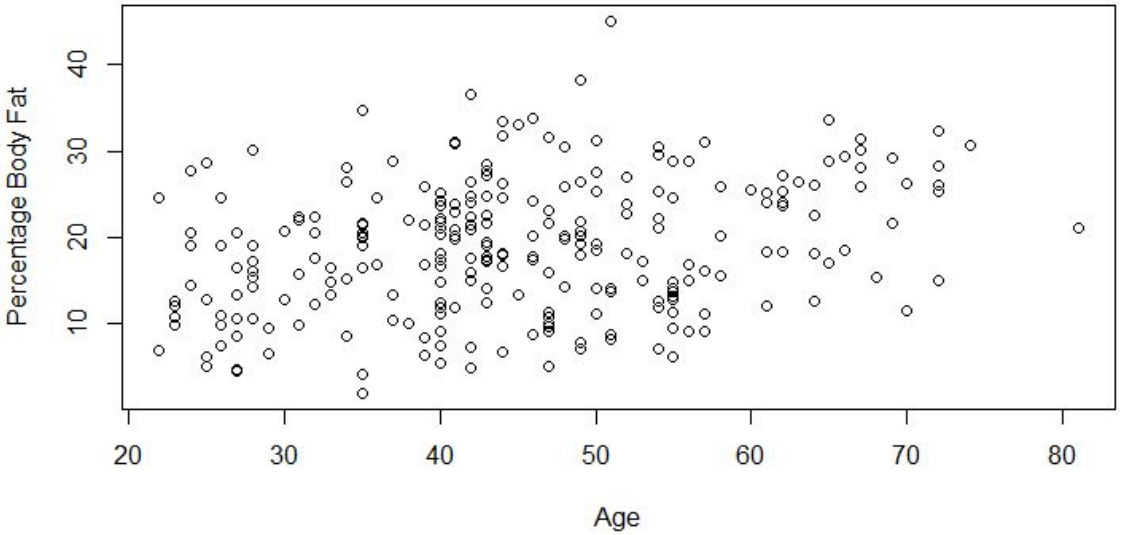


$R^2 = 66.21\%$

$R^2 = 53.00\%$

Both are seemingly significant

Initial Model Creation



$R^2 = 8.36\%$

Easy to use

Our Guess Model

We first used adiposity, abdomen, and age as predictors.

But we noticed that the slope for adiposity is negative and the p-value when performing an f-test removing adiposity is greater than 0.05

-> We decided to only use abdomen and age.

Body Fat Percentage =

Abdomen Circumference (cm)*.6 + Age*.1 - 40

Statistical Tests

1. **Test whether at least one of the two variables are important using F-test**

$$H_0 : \beta_1 = \beta_2 = 0, \quad H_1 : at\ least\ one\ \beta_j \neq 0$$

- P-value is less than 0.05, rejecting null hypothesis.

Statistical Tests

2. Test abdomen is important after accounting for age

$$H_0 : \beta_2 = 0, \quad H_1 : \beta_2 \neq 0$$

· Compare linear model only has age as a predictor with our model, using F-test

· We have P-value = 2.2e-16 <0.05, reject the null hypothesis.

· Same procedure for age

Statistical Tests

3. Test whether our model is as good as the model based on all variables, using F-test.

$$H_0 : \beta_1 = \beta_7 = 0, \quad H_1 : at\ least\ one\ \beta_j \neq 0$$

· The p-value is 4.212e-09 < 0.05,, so we reject the null hypothesis
· Conclude that the model based on age and abdomen is just as good as the model based on all variables.

Diagnostics

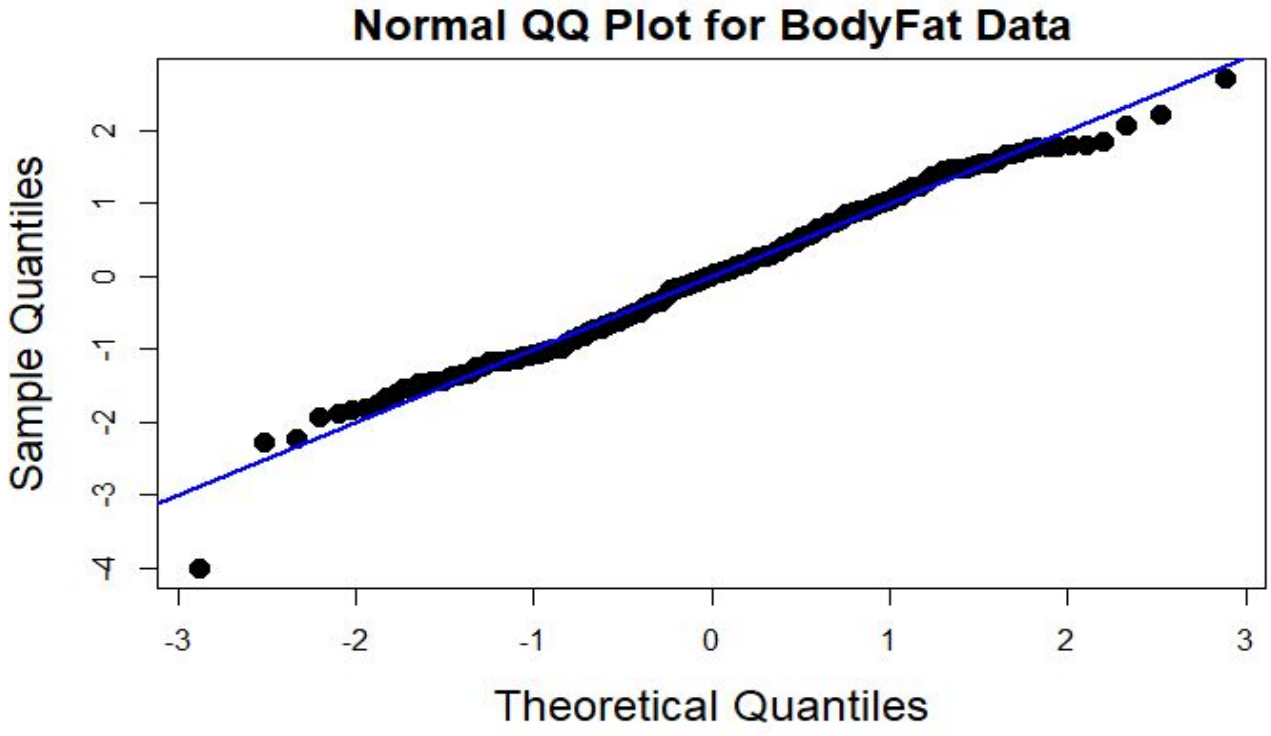
In order to make proper analyses, the regression must meet some key assumptions:

- Linearity
- Homoskedasticity
- Multivariate Normality
- No outliers
- Low level of multicollinearity

We find the Variance Inflation Factor (VIF) for each parameter is 1.062 which is much smaller than the threshold (5) to be considered variables with multicollinearity.

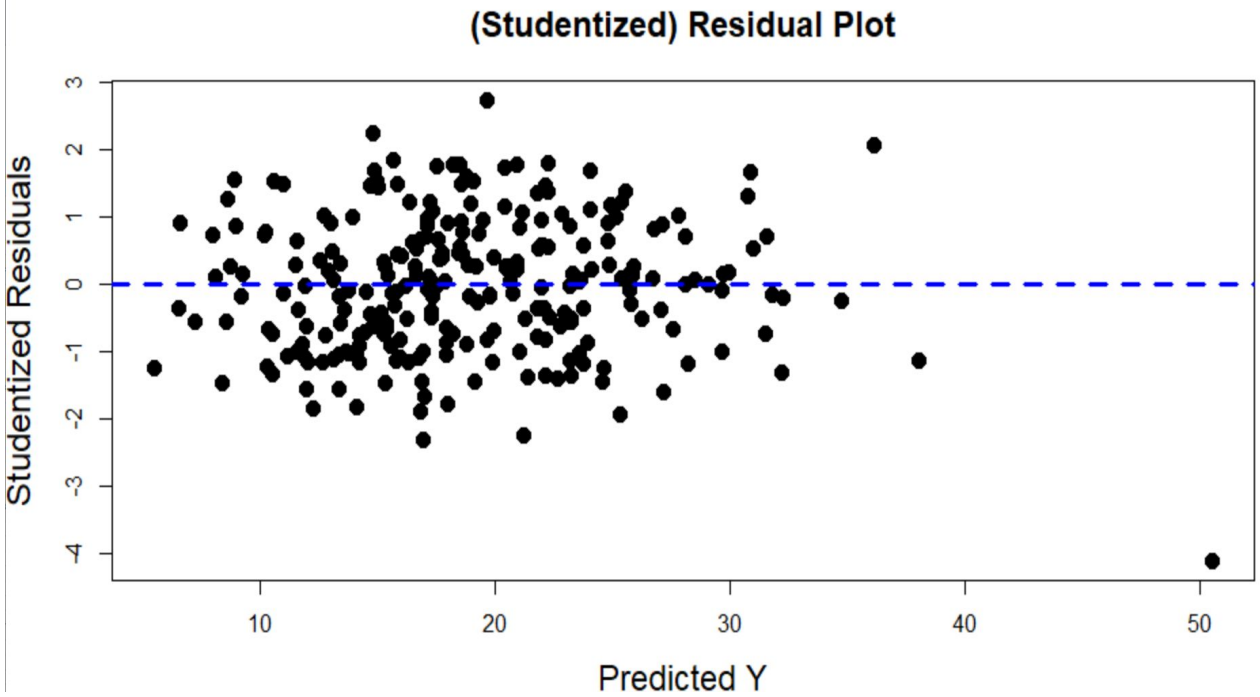
Diagnostics: QQ-plot

- Data hugs the 45 degree line and is normally distributed throughout
- We have enough evidence that our BodyFat dataset is normally distributed



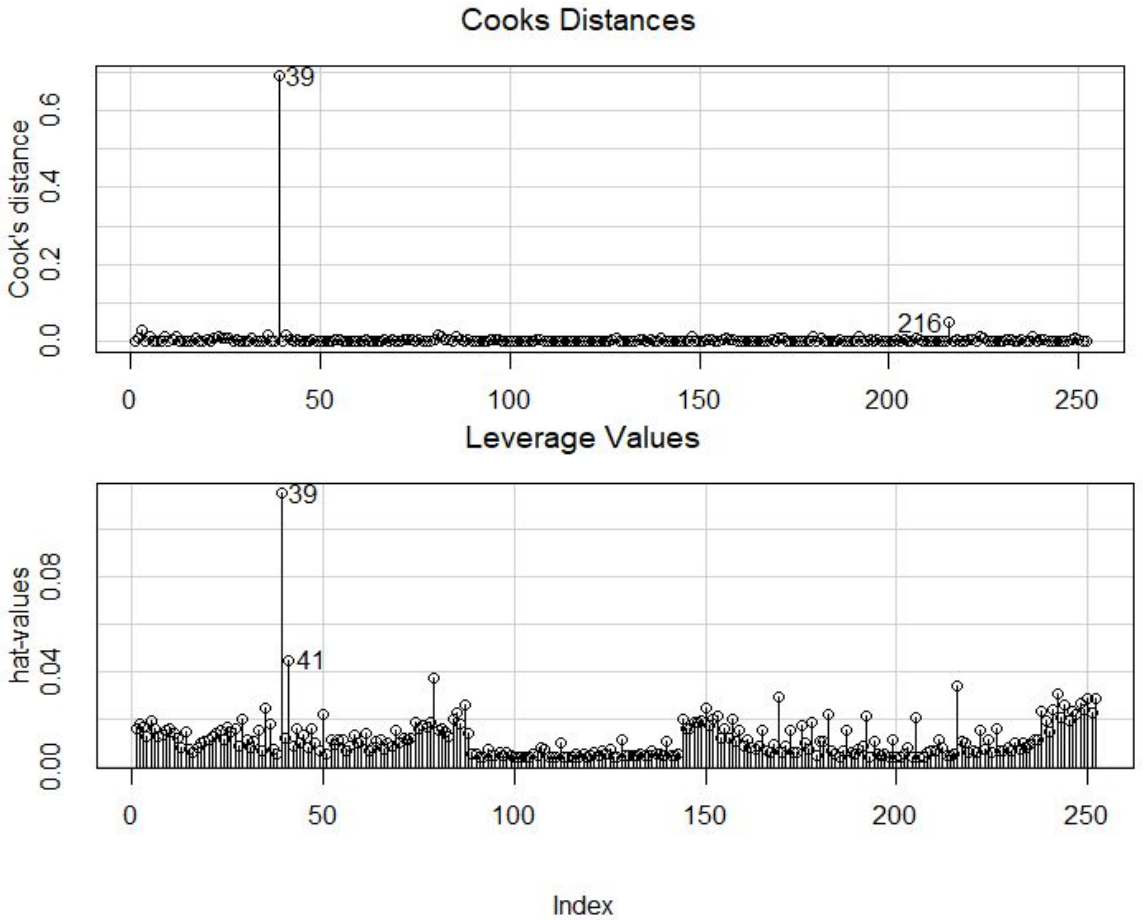
Diagnostics: Studentized Residual Plot

- No apparent non-linear trends
- We can confirm that body fat has a linear relationship with abdominal circumference and age
- No change in variability in y across the x-axis
- We confirm that the data are homoskedastic



Diagnostics

- Subject 39 produced a value of .7 a significant outlier
- Subject 39 is also an outlier among leverage values
- A larger Cook's Distance would provide stronger evidence that subject 39 is an influential point, however, removing the datapoint generated enough change for us to adjust our model



Strengths

- Easy to use
  - Abdomen circumference can be measured with tape measure
  - Everyone knows their age
- Relative accuracy
  - R^2 of 68.46% shows model to cover a significant portion of the variation in body fat percentage

Weaknesses

- Only works on men
  - Model cannot be generalized to women because the data set only included men
- Doesn’t work on children
  - Mean age of men was 45
  - Intercept of -40 means anyone very young or slim would likely under represent their true value