Bret Peterson

6/9/2019

# Predicting Body Fat

Percentage body fat (underwater) is often looked at by health professionals as an indicator of an individual's health. Although, it is a useful data point, calculating (underwater) percentage bodyfat is expensive and often inconvenient. As such, it would be useful to create a model to predict (underwater) percentage bodyfat. Using Ordinary Least Squares (OLS) and two different Neural Network Models; one basic and one more complex, I will analyze a dataset that contains various body circumference measurements. After determining which of the models is most accurate, I will use it to make a prediction for the (underwater) percentage bodyfat for a sample individual.

The first step was to run an Ordinary Least Squares (OLS) model with the Y-variable being Percent Bodyfat.

**Ordinary Least Squares (OLS)**

| Source | RSquare | RASE | Freq |
|---|---|---|---|
| Training Set | 0.7206 | 4.1208 | 151 |
| Validation Set | 0.8063 | 4.1121 | 50 |
| Test Set | 0.6657 | 5.1177 | 51 |

As can be seen in the table above, the results of the model were accurate on all the data and represented a 0.6657 R-Square on the Test Dataset.

After running an OLS model as a base model to compare other models to, the next model I decided to run was the default basic Neural Network model with 20 tours. The result is a slightly less accurate model with an R-Square value of 0.5987.

**Neural Network: Basic**

| Measures | Value |
|---|---|
| RSquare | 0.5986669 |
| RMSE | 5.6073595 |
| Mean Abs Dev | 4.4563325 |
| -LogLikelihood | 160.29394 |
| SSE | 1603.5665 |
| Sum Freq | 51 |

After running an OLS and a default Neural Network model, my next step was to run a more Complex Neural Network model, again using 20 tours. In doing so I had two layers and three nodes for all three activation functions.

**Neural Network: Complex**

| Measures | Value |
|---|---|
| Rsquare | 0.5009456 |
| RMSE | 6.2528766 |
| Mean Abs Dev | 4.6897406 |
| -LogLikelihood | 165.85099 |
| SSE | 1994.0218 |
| Sum Freq | 51 |

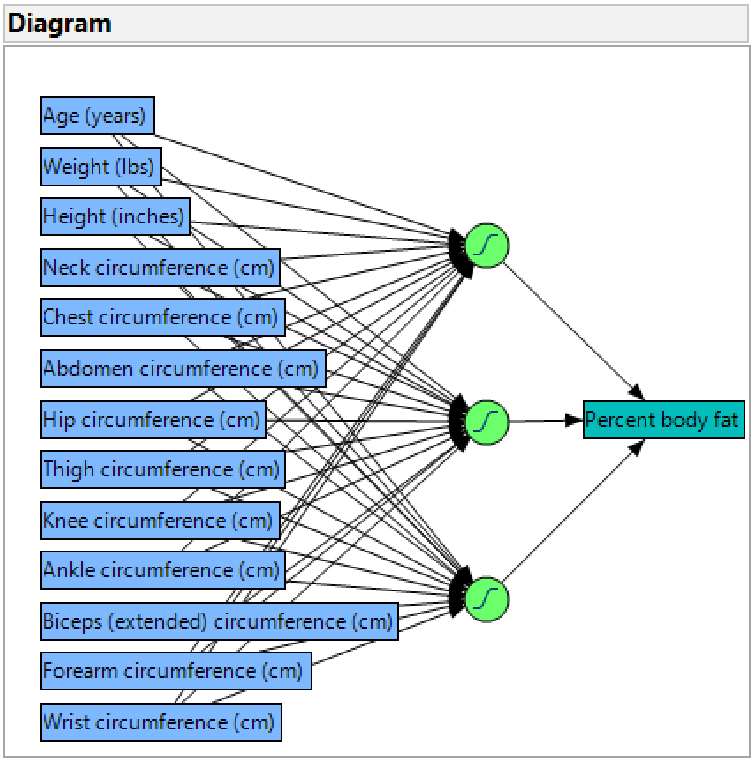
Again, the result is an accurate model with an R-square of 0.5009; but it is slightly less accurate than the Basic Neural Network Model and far less accurate than the OLS model.

Although, just from comparing the R-square values for the two models I can be confident which model is a stronger predictor, I will still do a model comparison to confirm.

**Model Comparison**

| Validation | Predictor | Creator | RSquare | RASE | AAE | Freq |
|---|---|---|---|---|---|---|
| Training | Predicted Percent body fat NN Basic | Neural | 0.7798 | 3.6582 | 2.9368 | 151 |
| Training | Predicted Percent body fat NN Complex | Neural | 0.7680 | 3.7557 | 2.9650 | 151 |
| Training | Pred Formula Percent body fat OLS | Fit Least Squares | 0.7206 | 4.1208 | 3.3320 | 151 |
| Validation | Predicted Percent body fat NN Basic | Neural | 0.8368 | 3.7735 | 3.0090 | 50 |
| Validation | Predicted Percent body fat NN Complex | Neural | 0.8298 | 3.8545 | 2.9443 | 50 |
| Validation | Pred Formula Percent body fat OLS | Fit Least Squares | 0.8063 | 4.1121 | 3.5172 | 50 |
| Test | Predicted Percent body fat NN Basic | Neural | 0.5987 | 5.6074 | 4.4563 | 51 |
| Test | Predicted Percent body fat NN Complex | Neural | 0.5009 | 6.2529 | 4.6897 | 51 |
| Test | Pred Formula Percent body fat OLS | Fit Least Squares | 0.6657 | 5.1177 | 4.0480 | 51 |

As can be seen in the table above, the “Ordinary Least Squares” model has a much larger R-square when looking at the “Test” data. Not only does this model have a higher R-Square as compared to the Basic and Complex Neural Network, it also has a lower AAE, again suggesting that it is a stronger model. Knowing that the OLS model is the most, it is next important to look at which variables had the biggest influence on the model. Outside of JMP, this is a very difficult task as it is hard to make a connection between variables as can be seen in the table below.

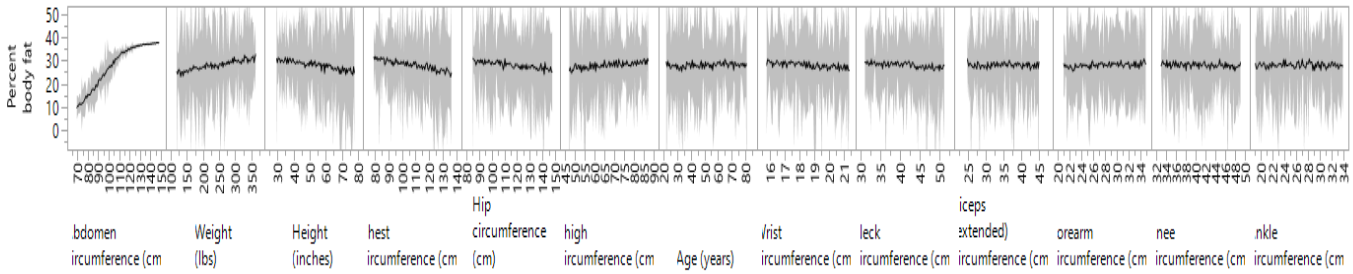



JMP offers an ability to better analyze which variables have the strongest impact on the model known as profilers. In looking at the Basic Neural Network "Variable Profilers" table below, it is obvious that "Abdomen Circumference" is the variable that has the highest impact on the percentage of bodyfat. This makes logical sense in that if you imagine an individual with a larger abdomen circumference, it is likely that they have extra bodyfat.

**Variable Profilers**

| Column | Main Effect | Total Effect |
|---|---|---|
| Abdomen circumference (cm) | 0.799 | 0.868 |
| Weight (lbs) | 0.028 | 0.047 |
| Height (inches) | 0.024 | 0.043 |
| Chest circumference (cm) | 0.026 | 0.043 |
| Hip circumference (cm) | 0.015 | 0.029 |
| Thigh circumference (cm) | 0.009 | 0.016 |
| Age (years) | 0.007 | 0.014 |
| Wrist circumference (cm) | 0.007 | 0.014 |
| Neck circumference (cm) | 0.005 | 0.01 |
| Biceps (extended) circumference (cm) | 0.003 | 0.006 |
| Forearm circumference (cm) | 0.003 | 0.005 |
| Knee circumference (cm) | 0.002 | 0.004 |
| Ankle circumference (cm) | 0.002 | 0.003 |

In analyzing the graphs of the variable profilers below, it is also clear that “Abdomen Circumference” has the highest correlation with bodyfat.




After determining the most accurate model, as well as analyze which variables had the strongest impact on the equation, I then used that model to predict the (underwater) percent bodyfat of a sample individual with the following measurements:

Age (years)=50

Weight (lbs)= 167;

Height (inches)= 67.75;

Neck circumference (cm)= 38.8

Chest circumference (cm)=100.4

Abdomen circumference (cm)=89

Hip circumference (cm)= 93.2

Thigh circumference (cm)= 57.0

Knee circumference (cm)=34.8

Ankle circumference (cm)= 20.6

Biceps (extended) circumference (cm)= 33.9

Forearm circumference (cm)= 28.3

Wrist circumference (cm)= 18.0

The most accurate model, the OLS model, predicted an (underwater) Percentage Bodyfat of 17.8139%. The Basic Neural Network Model predicted the (underwater) percentage bodyfat to be 20.862390809%. The Complex Neural Network Model, which was slightly less accurate, predicted a lower (underwater) percentage bodyfat of 18.804416815%.