**Predicting Body Fat**

**Susan Wiggins**

**Cases in Business Analytics BAN 525**

**Professor: Dr. Cetin Ciner**



**Table of Content:**

Introduction…………………………………………………………………………………………………………………………………….3-4

Analysis and Model Comparison………………………………………………………………………………………………………4-6

Interpretation………………………………………………………………………………………………………………………………….6-8

References……………………………………………………………………………………………………………………………………….9

Appendix A **(Ordinary Least Square Regression)**…………………………………………………………………………….10-11

Appendix B **(Neural Network one layer with three nodes with TanH Activation)**………………………….11

Appendix C **(Neural Network** **two layers with three nodes for all activation functions)**………………..12

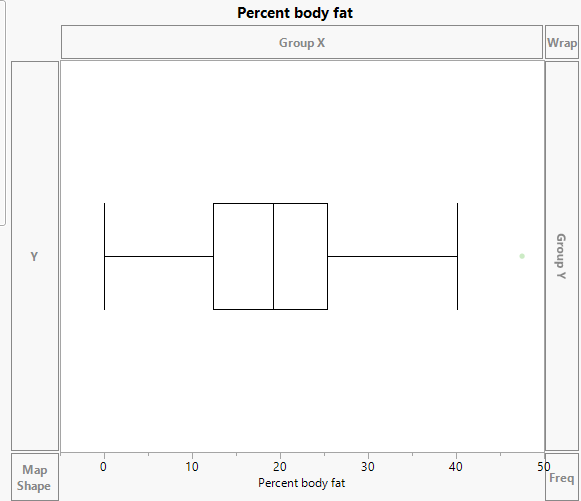
Appendix D **(Model Comparison )**…………………………………………………………………………………………………..12

Appendix E **(Neural Network Final Model including Profilers and Variable Importance Summary)**.13-15

Appendix F **(Sample Data Predication for percentage of body fat)**…………………………………………………15-16

**Introduction**

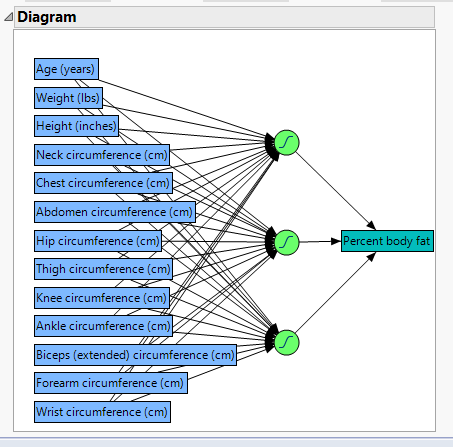
Health is a major concern of many people and health has been attested in booklets to be associated with percentage of body fat. An investigation of underwater percentage of body fat in men will be studied in this case as well as a predication of percentage of body fat based on given criterion set forth in this case. Elizabeth Quinn points out, “that underwater weighing is based upon Archimedes Principle which states that the buoyant force on a submerged object is equal to the weight of the fluid that is displayed by the object. You can use this principle to determine a person’s percentage of body fat because the density of fat mass and fat free mass are constant. Lean tissue, such as bone and muscle, is denser than water, and fat tissue is less than water. Because muscle sinks and fat floats, a person with more body fat will weigh less underwater and be more buoyant. Someone with more muscle, meanwhile, will weigh more underwater” (1). The dataset introduced has a count of 252 men. A healthy rate of body fat for men is between 8 and 19 percent. The spread of the data for the percentage of body fat was greatest between the 25 % Quartile at 12 and the 75% Quartile at 25. The median for the data occurred at the 50 % Quartile at 19 seen in the graph below. Also, one outlier occurs at row 5 and the data label for percentage of body fat is 47.5.



In addition, the dependent variable being measured in this study is percent body fat and the predicator candidates are Age, Weight, Height, Neck Circumference, Chest Circumference, Abdomen Circumference, Hip Circumference, Thigh Circumference, Knee Circumference, Ankle Circumference, Biceps Circumference, Forearm Circumference, and Wrist Circumference. All variables are continuous in this dataset meaning that the variables depict numerical values. Three methods will be addressed in this case ordinary linear regression, neural network assuming the default layer with one layer and three nodes with TanH as the activation function, and neural network assumption most complex model with two layers and three nodes involving all three activation functions. Neural Network which is the new method introduced in this case deals with complex relationships of inputs and outputs. Dr. Cetin Ciner points out that, “the specifics of a neural network involve input layer, one or more hidden layers, and an output layer. The hidden layer in a neural network will involve one or more nodes and a transformation will take place upon a linear combination. Neural Networks in JMP consists of three transformation functions which are TanH, Linear, and Gaussian” (11-12). The first Neural Network in this case involves the TanH which is similar to logistic function and the final model for Neural Network will involve all three transformation functions.

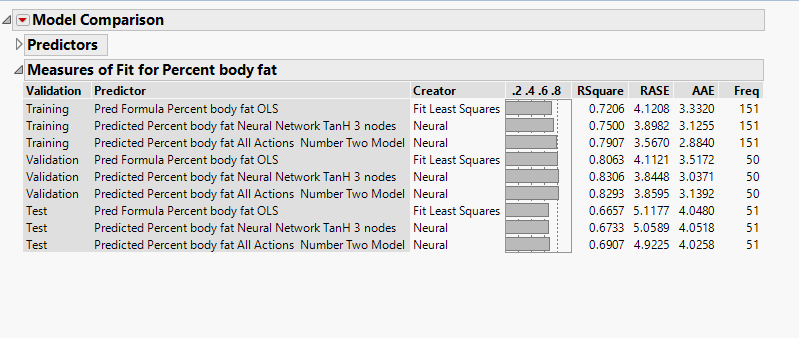
**Analysis and Model Comparison**

Ordinary Least Squares (OLS) a modeling technique most highly used in prediction methods can cause problems when Big data is being used. For instance, OLS modeling can cause large variances in the data, models random noise, and poor forecasting can exists. Therefore, other machine learning techniques need to be implemented such as Neural Networks. Neural Network (NN) is a complicated model to understand just through looking at the diagram below based on the first NN in the case. The green circular midways in the diagram represent the hidden layers. The left variables listed in blue box formations are the inputs and the right most blue box is the output in the Neural Network. The focal point is not as clear in the diagram and the profiler option becomes the point of clarification in this modeling technique. On the other hand, Neural Networks have high flexibility and have the ability to model very complex relations.



In order to reach the best results in the analysis one must depend on cross validation which will decrease the ability to interpret random noise. In cross validation, data is held out and the predication criterion is based on this interpretation. Estimates are then built on the data after the completion of each modeling affect occurs. In the case predication of percentage body fat, the cross validation is established on 60/20/20 split of the data with a random seed of 123. The final interpretation of the model will be based on the results of the test data due to the basis of new observations.

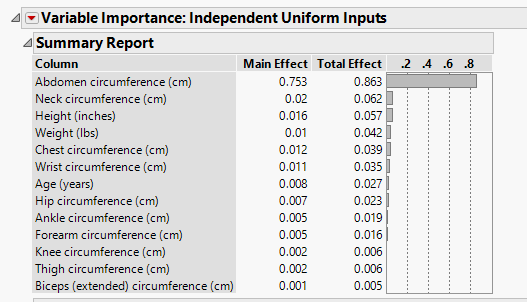
Next, the following modeling methods were performed Ordinary Least Squares, neural network assuming the default layer with one layer and three nodes with TanH as the activation function, and neural network assumption most complex model with two layers and three nodes involving all three activation functions. The results of each one of these models are listed in the appendix A, B, and C. A model comparison was run for analysis of best results of modeling predications and chart is listed below.



The model comparison revealed that Neural Network II based on the most complex model with two layers and three nodes involving all three activation functions was the best modeling technique used for predication of percentage of body fat. The explanation is base on the RSquare having the highest yield of 69.07% and the lowest value for root average square error at 4.9225. In addition, the average absolute error (AAE) was also the lowest in comparison to the models indicating Neural Network II best fit for the model. Therefore, Neural Network II explains 69.07% of variations in percentage of body fat and one can see RASE scores are much less than other models as well as lowest scores for AAE.

**Interpretation**

The variable importance is included in the predication of percentage of body fat and the following table gives us the results below.



The total effect takes into account the variations in the data and the Abdomen Circumference explains 86% of the variations in the percentage of body fat. The most important variable to look at is abdomen circumference. The second variable to study is Neck Circumference listed at a low percentage rate of 6.2% followed by the third most important variable listed as Height at 5.7% also low. The model prediction profiler reveals that Abdomen circumference has a positive relationship when associated with percentage of body fat, as a person abdomen circumference increases so does the percentage of body fat for men in this dataset. Next, variable to study in the model profiler is neck circumference which provides a negative relationship towards percentage of body fat, as neck circumference increase the percentage of body fat decreases. The third variable studied which makes a slight impact is height and height has a negative relationship, when height increase then percentage of body fat decreases. In addition, by studying the marginal model plots the analysis shows the greatest change in abdomen circumference in relation to percentage of body fat. A slight decrease in association with body fat is noted with the following two variables height and hip circumference whereas other variables within the case study remained stagnate.

A final step for the analysis of this case study was to predict a probability based on supplied parameters for field values in our study listed in Appendix F. The predication of body fat for a man reported based on the Neural Network II method chosen and under the parameters in Appendix F reported a 20.44% likelihood for the best predication in this method.

In conclusion, the selected model to predict percentage of body fat in men was Neural Network II. The highest variable for fluctuations in percentage of body fat is related to the abdomen circumference at 86 percent and has a positive relationship. The second most important variable associated with percentage of body fat is the neck circumference and has a negative relationship on percentage of body fat. The final variable studied in the Neural Network II is the height which only takes into account a small percentage at 5.7% and has a negative relationship in account that when the height increases the percentage of body fat decreases. Therefore, one could associate taller people having leaner body fat percentages. In addition, abdomen circumference which is greater would predicate a man to have a greater percentage of body fat. Overall, the variable which makes the greatest impact on percentage of body fat for men is abdomen circumference. Also, RSquare for modeling selection was the highest for Neural Network II at .6907. Therefore, Neural Network II explains 69.07% of variations in percentage of body fat. Finally, the predication of body fat for a man in this dataset reported based on the Neural Network II method chosen and under the parameters in Appendix F reported a 20.44% likelihood for the best predication in this method.

**Reference**

Ciner, Cetin. August 2019. Module 1. “Cross Validation and Predicting Gold Prices.” Slides 1-53. Retrieved from: <https://uncw.instructure.com/courses/24252/pages/module-1-lecture-materials-cross-validation-stepwise-regression-and-risk-factors-for-gold-prices?module_item_id=524495>

Ciner, Cetin. September 2019. Module 4. “Neural Networks and Customer Churn.” Slides 1-27. Retrieved from:

Fisher, Garth A., "Generalized body composition prediction equation for men using simple measurement techniques", K.W. Penrose, A.G. Nelson, A.G. Fisher, FACSM, Human Performance Research Center, Brigham Young University, Provo, Utah 84602 as listed in \_Medicine and Science in Sports and Exercise\_, vol. 17, no. 2, April 1985, p. 189.

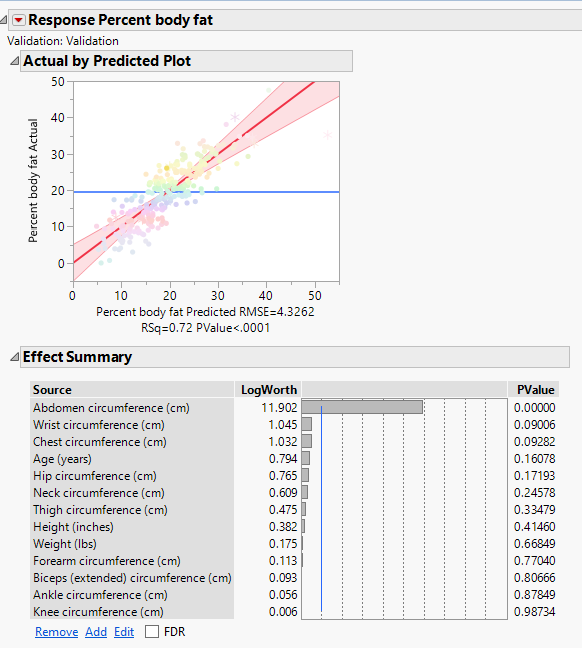
Madison, David. Getty Images. Website. Picture. <https://www.verywellfit.com/what-is-hydrostatic-underwater-weighing-3120276>

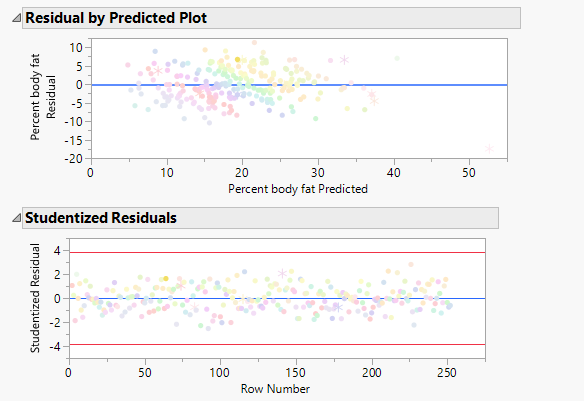
Quinn, Elizabeth. “How to Measure Body Fat with Hydrostatic Underwater Weighing.” Health and Safety. 13 Sept. 2019. Web. 26 Sept. 2019. <https://www.verywellfit.com/what-is-hydrostatic-underwater-weighing-3120276>

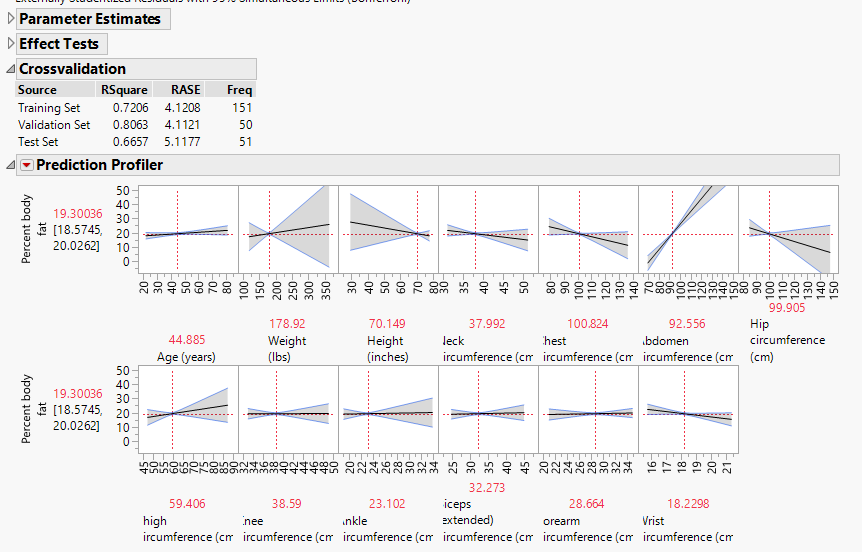
Website. <https://www.google.com/search?source=hp&ei=ydCMXZ20PM6AtgX--L6oBQ&q=body+fat+percentage&oq=body+fat&gs_l=psy-ab.1.0.0l10.891.2558..5200...0.0..0.73.531.8......0....1..gws-wiz.......0i131.VGDVuNUm-7o>

**Appendix A**

**Ordinary Least Squares Regression**

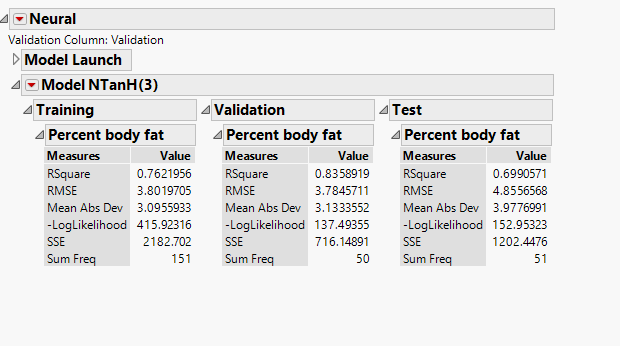






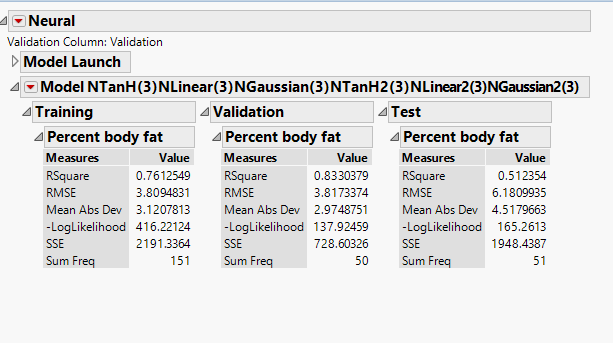
**Appendix B**

**Neural Network one layer with three nodes with TanH Activation**



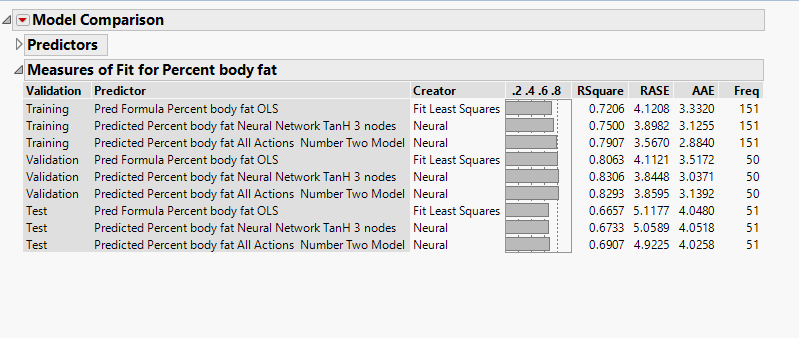
**Appendix C**

**Neural Network two layers with three nodes for all activation functions**



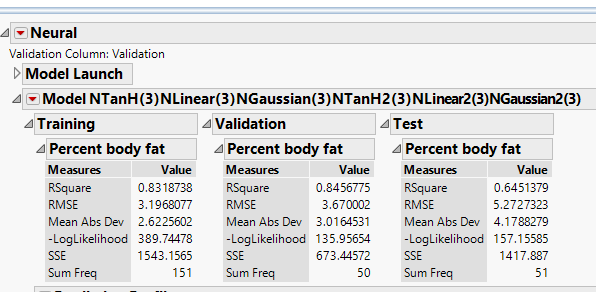
**Appendix D**

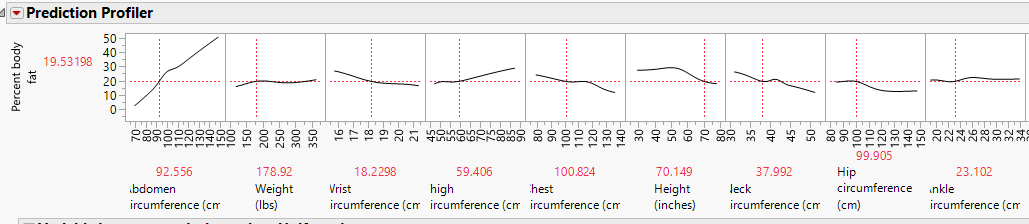
**Model Comparison**

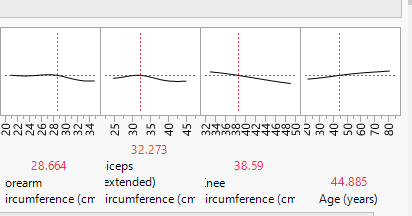


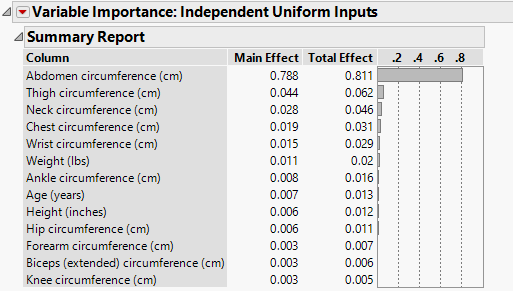
**Appendix E**

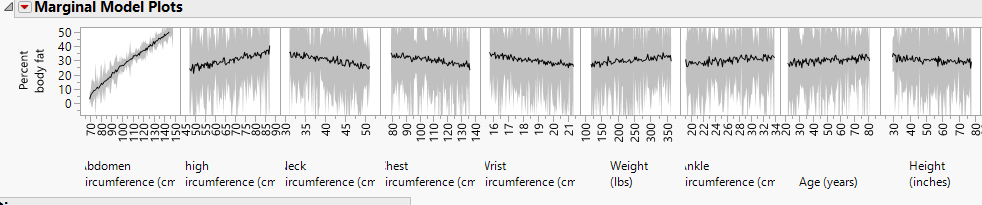
**Neural Network Final Model including Profilers and Variable Importance Summary**

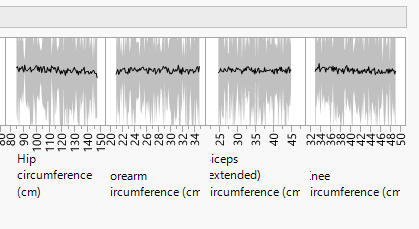


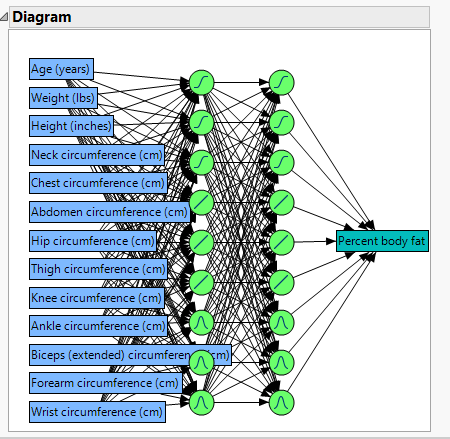












**Appendix F**

As a new sample, suppose you have a male with the following data, what is your best prediction for his percentage body fat?

|  |  |
| --- | --- |
| 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 |
| 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 |

