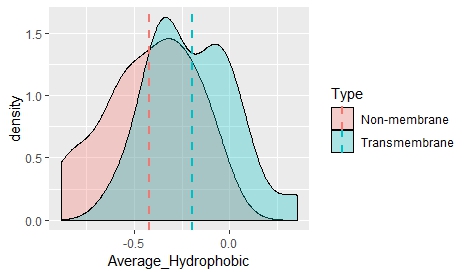
Capstone Project: Summary Statistics

Zach Young

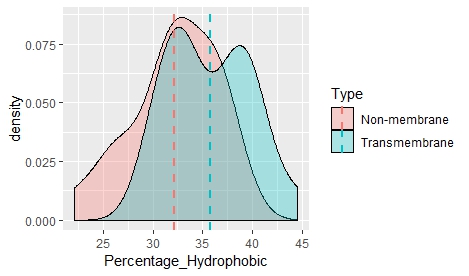
Introduction:

Knowing the chemical nature of biological membranes and their affinity towards hydrophobic molecules, I have come up with the hypothesis that transmembrane proteins should contain more amino acid residues of a hydrophobic nature than their counterpart non-membrane proteins. Based off this hypothesis, I have focused my statistical analysis on three key measurements. First, I have decided to find the average hydrophobic value for every protein I have chosen. This was accomplished by simply finding the mean hydrophobic value of all amino acids contained within each particular protein. If my hypothesis is correct, transmembrane proteins will contain more amino acids with higher hydrophobic values and thus have a higher overall average hydrophobic value than non-membrane proteins. Secondly, I will look at the percentage of all hydrophobic amino acids within a protein compared to the total number of amino acids in the protein. As with the average hydrophobic value, I’m predicting transmembrane proteins will have a larger amount of hydrophobic amino acids and in turn a higher percentage of hydrophobic amino acids. Finally, I will examine the percentage of each individual amino acid within a protein and try to ascertain any significance between the two types of proteins.

Average Hydrophobic Values:

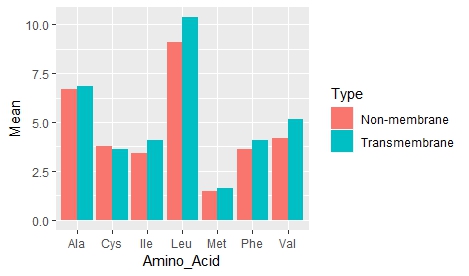
 Having calculated the average hydrophobic value for every protein chosen for analysis, I then calculated the mean, variance and standard deviation for the entire group of each type of protein, non-membrane vs. transmembrane. The mean average hydrophobic value for transmembrane proteins was -0.194, with a variance of 0.052 and a standard deviation of 0.228. The mean average hydrophobic value for non-membrane proteins was -0.422, with a variance of 0.057 and a standard deviation of 0.240. These numbers show that transmembrane proteins do seem to have a higher overall hydrophobic value on average than non-membrane proteins. The density plot to the right also seems to back the hypothesis.

Percentage of Hydrophobic Amino Acids:

After calculating the percentage of hydrophobic amino acids in every protein under analysis, I then found the mean, variance and standard deviation for each group of proteins. For transmembrane proteins, the mean percentage of hydrophobic amino acids was 35.781 with a variance of 17.092 and standard deviation of 4.134. For non-membrane proteins, the mean percentage of hydrophobic amino acids was 32.198 with a variance of 19.195 and standard deviation of 4.381. Like the average hydrophobic values, these values indicate that on average there are more hydrophobic amino acids in transmembrane proteins as opposed to non-membrane proteins.

Percentage of Individual Amino Acids:

Finally, I have decided to look at the percentage of each individual amino acid within all proteins to determine whether or not one particular amino acid is over- or under-represented within either group of proteins. The histogram below shows the mean of the hydrophobic amino acids between both types of proteins. Having examined this chart, it appears that the amino acids Leu and Val may have significance in determining non-membrane vs transmembrane. I have decided to perform a t-test on both these amino acids to ascertain the level of significance.



T-test Values:

To determine if there is significance between the above characteristics, I have decided to perform t-tests on four separate values. Performing t-tests on the average hydrophobic value, percentage of hydrophobic amino acids, percentage of Leu and percentage of Val all resulted in t-values greater than 2 and for average hydrophobic value, the t-value was over 3. This seems to indicate that there is significant difference in all these values between transmembrane and non-membrane proteins. As a result, I will use these four characteristics in my model to accurately predict transmembrane vs. non-membrane proteins.