STAT 630

Group 3

**Screening Tool Questionnaire Chart**

**Explanation:**

Our group created this questionnaire to simplify the process of predicting patients with a high-risk of Chronic Kidney Disease. The initial steps before formulating the essential questions necessary to comprehensively represent the results of our logistical regression model, we had to identify the most significant variables.

After conducting the research about Chronic Kidney Disease, we concluded that the status of diabetes, hypertension, cardiovascular disease, congestive heart failure and the age group of an individual contributed the most to CKD. We used a binary scale for each variable, denoting a value of 1 when a variable was relevant to a patient and a 0 when it was not. We decided to use a binary scale to mirror the results obtained from our logistic regression model, because we wanted to compare the two results to check the simplified screening tool’s accuracy. After assigning binary values to patients for each variable, we summated the rows. The summations were then independently compared with a threshold value that we set as a standard to decide if a patient had a high-risk of CKD. The threshold value we chose was 1.8, which we decided by weighing out each variable. We assumed that a patient could be classified as high-risk, if 2 out of the 5 variables pertained to him. If the summation of a patient’s row was greater than 1.8, we assigned a 1, and we assigned a 0 for patients with less than or equal to sums to 1.8. However, a slight obstacle we had to navigate through was the weightage assigned to specific age groups. We quickly realized that more elderly age groups contributed more to CKD than younger age groups. Our solution was assigning values of .2 for ages 20 to 35, .5 for ages 36 to 55, .8 for ages 56 to 69 and 1 for ages 70 and older. This formulation made it possible to subset the age variable and put more emphasis on the 70 and older group. The age variable is how we arrived at a threshold value of 1.8. We found the average of the 56 to 69 and 70 and up values made sure the 70 and up age group would be able to carry a row above the threshold value.

Finally, we ran a correlation function in excel between our logistical regression model results and our screening tool results to visualize the screening tool’s accuracy. There was a 0.7 correlation between the two various predictive models. We were satisfied with a 0.7 correlation because the screening tool acted as a simplified model of our logistical regression model.