

MATH205

Jake Schichko, Calvin Banning, Jacob Petit de Mange

Flow of R Code for Final Project:

Our R code starts off with selecting a file and then immediately goes into reading the file and breaking up the values that we will use later such as the length, what characters are given to us as well as other data such as the quartile values that we will use to determine what kind of test we are using. Our code then goes into more detail of basic information gathering before we decide which tests we are going to use, this includes doing things such as the $1.5 \times \text{IQR}$ rule to determine if we have any outliers as well as using the shapiro test to determine if our data is approximately normal. Then we start to get into determining what types of tests we are going to use, using all of the information we have gathered as well as the two Characters given to use, displaying details such as “center”. We decided it would be best to start as broad as possible with the CSK value. This is because if we have the CSK = ‘K’ we know the test must be a proportions test, and if we have the CSK = ‘S’, we know that the test must be an F test so that knocks those two out of the way right at the start. We then added more description to the tests such as for the proportions test we also checked to make sure that n1 and n2 were both greater than 10 and if they are we can proceed to carry out with the test and print our descriptions, otherwise we must abort the test. But our main goal was to start as broad as possible, so we know what tests we certainly can not do. Later when we got to if CSK == ‘C’, which contained four of the 6 tests we were to perform, we again tried to start as broad as possible and then break down by looking at DI. We determined if DI == ‘I’ It had to be either the pooled two sample test or the regular 2 sample test and if DI == ‘D’ it needed to be the paired t or sign test. Again, once we figured out what DI was equal too we broke it down even further to get to which of those 2 tests to use. For example, when DI == ‘I’, if our variable equalVar was true, we know we had to use to the pool two sample test, and if not we used the two sample t test, and if for either of these, the normalX and normalY were less than 0.05 we had to abort. So here we broke 4 choices down into 2 very easily and then after that was done we could break them down even further by looking at only one or two variables.