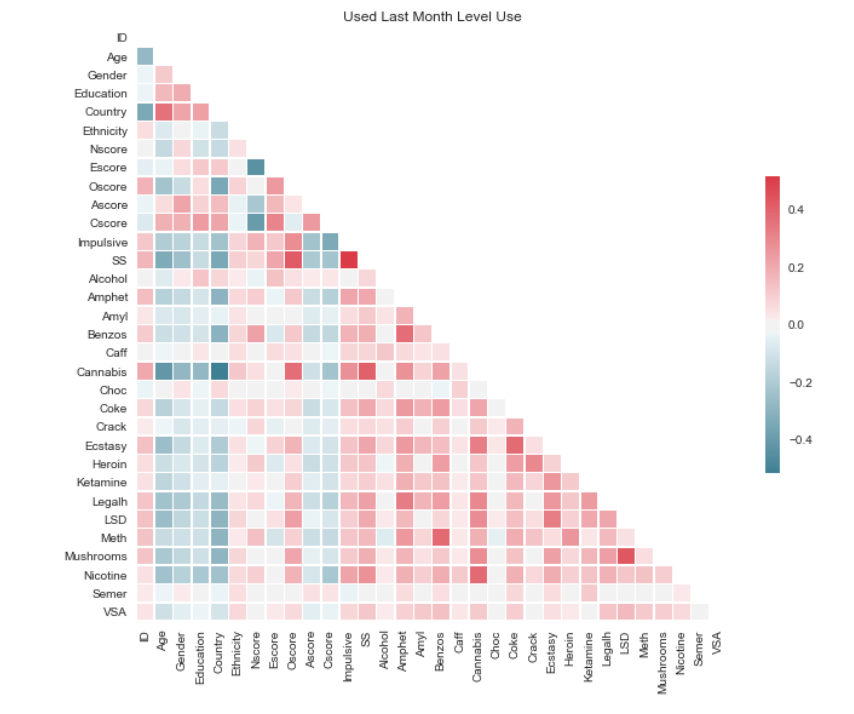
Apply Inferential Statistics to Capstone Project 1

Alissa Mezzacappa

For drug use in the last month, there appears to be a strong positive correlation between cannabis and sensation seeking as seen in the figure below. Is the correlation of Cannabis to Sensation Seeking significant?



Null Hypothesis: population of cannabis users in this class that DO NOT have SS attribute = population of cannabis users in this class that DO have SS attribute

Alternative Hypothesis: population of cannabis users in this class that DO NOT have SS attribute != population of cannabis users in this class that DO have SS attribute

For this attribute in the data a zero value represents a neutral value of sensation seeking. Positive indicates sensation seeking, negative indicates non sensation seeking. We break the data into these two classes with the code below:

C3t=data3[(data3.Cannabis)==1]

p\_prop= C3t.SS.mean()

p\_c=C3t.SS.count()

p0\_prop=C3t[C3t.SS<=0].SS.mean()

p1\_prop=C3t[C3t.SS>0].SS.mean()

p0\_c=C3t[C3t.SS<0].SS.count()

p1\_c=C3t[C3t.SS>0].SS.count()

From the above code we see for the subset of Cannabis users, the sample mean for sensation Seeking is 0.4541 with a sample size of 788. Sample mean for non sensation and sensation seekers is -0.6195 and 0.8217, with sample size of 201 and 587 respectively.

If the null hypothesis is true the true mean of the sample will be zero.

So I ask, "Is the true mean of the population 0 rather than 0.454?"

mu=0

SE=C3t.SS.std()/p\_c\*\*0.5

Z=(p\_prop-mu)/SE

Z= 15.37

This Z value falls outside of the range for Confidence Level=95% (-1.96,1.96). Therefor 0 unlikely to be the true population mean, and we can reject the Null Hypothesis in favor of the Alternative Hypothesis verifying the statistical significance of the correlation.