Kenny Waite DSC 530 Final Term Project

Does a person's adjusted gross income play a factor in whether they will have their taxes done by a volunteer in North Carolina? That is the central question I am attempting to answer with this project. My hypothesis is yes, people in lower income tax groupings take advantage of volunteer tax return services more than people in higher income tax groupings. I also wanted to look at other variables to see if there were any trends in volunteer prepared returns and the people who are more likely to use this service including:

- The number of dependents for tax filer
- City
- Farm Returns
- Students

My exploratory data analysis (EDA) resulted in a couple of different findings from the various exploratory testing performed on the data.

- All five of my variables mentioned above skewed left towards zero. This is why I chose to use Spearman's Rank correlation test statistic instead of Pearson's Correlation test.
- The CDF showed that the most common values are are between 20 and 100. This means the most common amount of volunteer tax returns by zip code is between 20 and 100. Also, the CDF showed me that about 10% of volunteer tax returns are between 0-10 and about 90% are less than 300 volunteer filed tax returns.
- Plotting the variables on a scatter plot highlighted that there was negative linear relationship between the number of volunteer tax returns filed and the income grouping. This calculation tells us that as income increases, the number of volunteer tax returns decrease.
- A second plot of the number student filed tax returns and the number of volunteer filed tax returns resulted in a positive linear relationship. this calculation tells us that as number student tax return filers increases, the number of volunteerÂ tax returns increases.

The last, and maybe most important finding from my EDA was that there was a stistcically significant relationship between the income grouping and the number of volunteer tax returns filed by zip code. When performing a regression analysis on the

effect of income grouping on volunteer tax return, the R squared resulted in 0.11 with a p-value less than 0.05. This indicates that the model is statistically significant with the income grouping predicting 11% of variance in the amount of volunteer tax returns filed.

Link to final assignment - https://github.com/kwwaite1129/Example-Project-Work/tree/master/Volunteer%20Tax%20Return%20Project