This assignment seemed much more straightforward than the “widgets” assignment. I created a script in Python along the specifications provided, though there are still a few questions. If I were to move code like this into production, I would want to know more about the nature of the input files.

Will the script always be examining populations or any other numbers? For some overly constrained values, Benford’s assertion would be false (telephone numbers) additionally, some distributions are too tight to achieve Benford’s assetion (adult human height’s in meters).

Entry errors are to be expected, but I would also like to check a few more datasets for potentially erroneous parsing to give me some ideas to make the “first digit grabbing” more robust.

The .dms file format was odd in that it seems to be a rarely used file extension today, however it processed fine in plain .txt format both in PowerShell (which I used to check line counts and a few other things) and in Python.

One of the larger issues come from the definition of “close to 30%.” I used 27% to 33% allowing 10% (of 30%) on either side, but I chose this arbitrarily as a “nice” round number. It’s possible statisticians have developed more precise tolerances, but reducing to just “true” or “false” based on this arbitrary choice could certainly be improved.

The assignment mentions returning an object, I would like to know the downstream use cases for the object to see what other improvements could be made.