Suraj Shah

Analysis of Probability Project

* Created script ExcelProb.py that begins with filename and path to the file. Can be universal in the file chosen, making it easily adaptable
* First function was called pretest, which creates a new sheet with only the pretest data
  + Creates a dataframe and extracts only all data that has pretest stage
* Second function is called grades, and creates a new sheet with only the score data
  + locates columns that have scores, renames the columns in the dataframe, and writes to a new sheet
* Third function, allgrades, only copies the data from those who completed all the questions
  + replaces empty cells with an NaN and drops all rows that have an NaN
  + writes to a new sheet
* Fourth function, correct, finds all rows that do not equal 1 and replaces with empty string
  + Writes to a new sheet, that stores only correct scores
* Fifth function, conditional, uses data by importing and creating a dataframe from the correct worksheet
  + Replaces all values with integers and finds where column i and column i – 1 are the correct score
  + Uses range and counts backwards, because that was the most effective way
  + Writes to new sheet
* Sixth function, cumulative, uses three dataframes, and uses the data from Correct, Conditionl, and an empty dataframe
  + locates all values in which Conditional dataframe has the correct score, and the i -th term has the correct score
  + Writes this data to a new sheet
* Seventh function is the statistics function, which is where all the calculations are
  + It uses 4 dataframes from data from the Correct, AllGrades, Conditional, and Cumulative sheets
  + It counts the total number of grades and also the sum of the correctGrades
  + It makes several lists, and creates new dataframes, each dataframe representing a staticistical anaylsis
  + For example, the first new dataframe calculates the probability of each answer being correct
  + The next two dataframes calculate conditional probability
  + The next two dataframes calculate cumulative probability
  + The next two dataframes verify the answered calculated above
  + Then the last dataframe is created through a list that verifies the cumulative data