

Stevan Maksimovic

*How an Author Affects His or Her Numerical Style of Writing*

What is the relationship between an author and his/her numerical style of writing? Using Python v2.7.2, a program was made that reads the first 100000 characters of a book in a text file, calculates the average number of words per sentence, the average number of characters per word, the average number of characters per paragraph, and the average number of characters per quotation, and counts the total number of punctuations and the total number of characters in quotations. With this program, data was extracted for all of these variables for 4 authors with 8 books each to perform data and statistical analysis. In data and statistical analyses, the mean and standard deviations were calculated as the key components for further analysis. Then, the Student's t-test was used for every combination of authors for each variable to find out if the means of the samples are significantly different, and normal probability distribution was used to find out how much more likely data points would be in one data set than another. After the completion of this statistical analysis, it was determined who the author most likely was of every book in the data set and six books outside the data set. The success rate of the author identification program was 78% for the books in the data set, and 67% for the additional books, both higher than the success rate of 25% for a random choice.