Name(s)

Ryan M. Anielski

Project Number

J1602

Project Title

To Type or Not to Type: The Probability of Words Appearing in Randomly Generated Letter Sequences

Abstract

Objectives/Goals
To determine the number of three, four, five and six letter words appearing in randomly generated letter sequences using both the 26 letters of the alphabet and the letter distribution from a standard Scrabble set.

Methods/Materials
Computer programs were written using "Phrogram" software to generate 10 sets of 10,000 random letters using the 26 letters of the alphabet, and 10 sets of 10,000 random letters using the letter distribution from a standard Scrabble game. The sets were checked and triple checked for 3, 4, 5 and 6 letter words, which were highlighted, verified using a Websters English Dictionary, and counted.

Results
The results of Test 1 (using 26 letters of the alphabet) produced approximately 50% less words than the expected outcome based on the statistical probability of finding words. Test 2 produced approximately twice the amount of 3, 4 and 5 letter words as Test 1, and 6 times more 6 letter words. Test 2 produced close to the statistical probability of 3,4 and 5 letter words and more than six times the expected probability of 6 letter words.

Conclusions/Discussion
The amount of words appearing in sequences of random letters is far less than the statistical probability would indicate. Most words that appear are 3 letters long, with the number of words decreasing as they get longer. When the Scrabble letters are used, far more words appear and their numbers generally are close to the statistical probabilities.

Summary Statement

This project compared words appearing in random letter sequences to statistical probabilities of words appearing.

Help Received

Math teacher at school helped explain statistical probability, uncle helped locate programming software, parents helped edit paper

Ap2/09