Letter Frequency Distribution Investigation

Name: _____ Date: _____

This investigation will be done to determine a statistical break-down of the frequency of letter use in the English language. We will analyse some random texts, make appropriate bar (column) charts of frequency and percent frequency. We will also compare the results with a secondary source of a very large sample size.

The Sample Text

"My argument was based mostly on anecdotal evidence. I do have an advance degree in psychology and my thinking was soundly based on accepted psychological theory. But it was hardly based on comprehensive research. Instead, it came exclusively from my own personal experience of playing with my own children.

The truth is, I'm not really a gamer. In fact, I didn't even play many video games as a kid. I had an Atari, but never a Nintendo. When I went to my friend's houses, I didn't find their gaming set-ups particularly stimulating. This is probably because, without having time to practice, I never got good enough to win. I was disinterested, bored, and even a little critical.

My fascination with video games began when I bought a console for my own kids. On Ebay, I won a Wii bundled with a bunch of games. Together, the three of us played *Mario Kart, Rayman Origins, and Super Smash Brothers*, etc."

Activity

Create a rough frequency table for each of the letters in the alphabet (a → z). Use a tally method to count the frequency of each letter from the sample paragraph. From the rough table create a table in a spreadsheet and then create two graphs. One bar chart shows the letter and the frequency. The other bar chart shows the letter and percent frequency. Create a Word document with headers, titles, page numbers and a description of the task. Add the graphs to your Word document.

Comparing Your Results

Copy and paste the English relative frequency chart (http://en.wikipedia.org/wiki/Letter_frequency) into your word document with your analysis. This chart is in a form called (relative frequency). Simply multiply the numbers by 100% to get the percentage. Compare a few of your values from your sample document with the "accepted" distribution. Are they similar? Comment on the similarities or differences and why these might occur.