Zipf Distribution

George Zipf popularized an ideaZipfs Lawthat approximates populations of cities, distribution of money in counties, and how frequently words are used. Nobel Prize-winning columnist Paul Krugmans wrote of Zipfs Law that

> *the usual complaint about economic theory is that our models are oversimplified that they offer excessively neat views of complex, messy reality. [In the case of Zipfs law] the reverse is true: we have complex, messy models, yet reality is startlingly neat and simple.*

A Zipfian Distribution: How Often Words Appear

Zipf Distribution

- A Zipfian distribution is a type of power law. A power law occurs when one event varies as a power of another. One application of Zipfs law states that in texts of natural language (e.g., books), each word is used twice as often as the next most commonly occuring word.
- The graph below applies the rule to word usage in 29 UK books below. The occurred 225,300 uses, and was the most commonly used word. Note that the graph is interactive; you can press the play with this data link to edit, embed, and share your own version.

Zipf Distribution

Evaluating Power Laws

- We can test for a power law by plotting frequency (y-axis) against rank (x-axis) on a double log axis. Then check for a straight line.
- The graph below shows three attempts to fit a power law function to datasets. The plot on the left is a good fit. The plot in the middle is a decent fit.
- The plot on the right is not a good fit.

Zipf Distribution

Evaluating Zipfian Distributions For City Populations
Another application of Zipfs law is for populations. Weve used ggplot2 to graph the population of cities (y-axis) and the rank of each city. In this dataset, New York has the highest population and is ranked first.

Zipf Distribution

GDP Of Nations

We are approaching a Zipfians distribution for country GDP vs rank.

Zipf Distribution

Evaluating Power Laws For Many Datasets

Researchers use power laws to determine how much inftrasture a city needs, examine the number of gas stations required in a city, and much more.