Log-Log Plots
Paul Schuette, Meredith College

Many elementary statistics courses now cover linear regression. A natural extension of linear regression includes simple data transformations. Logarithmic transformations are natural to consider in this context. If one can obtain the relationship $\log(y) = a\log(x) + b$, it follows that $y = 10^b x^a = Ax^a$. Thus, there is a basic relationship between log-log plots and power distributions.

Historically, Pareto (1895) was the first to observe this relationship in conjunction with his investigations into the distribution of wealth. Students may investigate more recent data at http://www.forbes.com/lists/. A plot of rank versus wealth for the top 75 wealthiest individuals yields:

Scatter Plot

Scatter Plot

logworth = -0.511logrank + 1.77; r^2 = 0.96

It may be noted that for the log-log plot $r \approx -0.98$, showing a rather strong linear relationship. If fact, Pareto thought he had discovered a law of economics, analogous to the laws of physics.

Auerbach (1913) and Lotka (1924) observed similar relationships for cities. Using Census data (http://www.census.gov/statab/www/), a log-log plot of rank versus population for 75 cities shows

Scatter Plot

logpop = -0.728logrank + 6.75; r^2 = 0.99

Once again, a strong linear relationship ($r \approx -0.9926$), may be observed.

Amazingly enough, one can see a similar picture using word frequencies from James Joyce's *Ulysses*. In linguistics, this is known as Zipf's Law.

Scatter Plot

logfreq = -1.02logrank + 4.47; r^2 = 1.00

Finally, one can even show that such relationships hold for scholarly productivity (Lotka, 1926). Jerry Grossman's Edös Number Page http://www.oakland.edu/enp/papers.html contains information about the number of papers per mathematician. If we consider those who have written 39 or fewer papers indexed in Mathematical Reviews, we obtain

Scatter Plot

logauthors = -1.54logpaper + 5.26; r^2 = 1.00

Other examples may be found in biology (area/species relationships), insurance, economics, libraries (Bradford's Law), and the popularity of web pages. Many have speculated that some general phenomena can explain these distributions, but Bruce Hill and others have shown that with the right hypotheses, limit theorems can be proven.