Re-expressing data

- Goals: Make the distribution of a variable more symmetric; make the spread of several groups more alike.

- The "Ladder of Powers"

Transformations: Logarithm Facts

- if $\log(x) = y$, $e^y = x$

- $\log(xy) = \log(x) + \log(y)$

- $\log(x/y) = \log(x) - \log(y)$

- "order preserving property"

- $\log(\sum X_i) \neq \sum \log(X_i)$; thus the mean of $\log(X)$ is not equal to the logarithm of the mean of X

Rationale for using log transformation

- commonly used in analyzing environmental data; shown to be adequate on both physical and empirical bases (Ott, 1995)

- positive (right skew) common in measurement data

- compresses high values, pulls in outliers, achieves symmetry when *group with the larger average also has larger spread*

- *Interpretation*: For a log-transformed dataset with 2 groups X and Y, we can calculate mean[log(Y)]-mean[log(X)] and make an inference *on the original scale of measurement*, that is, inference about the ratio median(Y)/median(X)

Two sample t-analysis: Cloud Seeding

	$\overline{log(X)}$	$s_{log(X)}$	n
unseeded	3.99	1.64	26
seeded	5.13	1.60	26

- Does cloud seeding have an effect on rainfall?

- What is the magnitude of the effect?