The Natural Log Transformation

We will use $\log(Y)$ to denote the natural logarithm of a number Y. (Note that the natural log button on your calculator is probably labeled $\ln$.)

The natural logarithm is related to the number $e = 2.718281....$ The number e is an irrational number like $\pi = 3.141592....$ (Irrational numbers are nonterminating, nonrepeating decimals that cannot be expressed as the ratio of two integers.)

$\log(Y)$ is defined as the power that you would need to raise e to in order to end up with Y; i.e., $e^{\log(Y)} = Y$.

For example, $\log(7) = 1.945910149$ because $e^{1.945910149} = 7$.

(On many calculators e^x is $\exp(x)$.)

If a set of data values is skewed to the right, taking the natural log of each data value can result in a data set that is roughly symmetric and often roughly normal.

(See Display 3.8 on page 69 of your text.)

The t-tools work best with symmetric, roughly normal populations. Thus it is often better to analyze the natural logarithms of the data values rather than the raw data when the raw data are right skewed.

If two right-skewed populations have different means and different standard deviations, taking logarithms can result in two roughly normal populations with approximately equal standard deviations.

(See Display 3.8 on page 69 of your text.)

Two-sample t-tests and confidence intervals for differences between two population means are most reliable if the two populations are roughly normal with approximately equal standard deviations. Thus it is often better to analyze the natural logarithms of the data values rather than the raw data when two populations appear to be right skewed with different standard deviations.

Y	2	4	8	16	32
$Z = \log(Y)$	0.69	1.39	2.08	2.77	3.47

Note that

$$\log[\text{Mean}(Y)] = \log\left(\frac{2 + 4 + 8 + 16 + 32}{5}\right) = \log(12.4) \approx 2.52$$

<div align="center">is NOT equal to</div>

$$\begin{aligned}\text{Mean}[\log(Y)] &= \frac{\log(2) + \log(4) + \log(8) + \log(16) + \log(32)}{5} \\ &\approx \frac{0.69 + 1.39 + 2.08 + 2.77 + 3.47}{5} = 2.08.\end{aligned}$$

However

$$\log[\text{Median}(Y)] = \log(8) \approx 2.08 = \text{Median}[\log(Y)].$$

For symmetric populations

$$\text{Mean}[\log(Y)] = \text{Median}[\log(Y)] = \log[\text{Median}(Y)].$$

For symmetric populations the sample mean of $\log(Y)$ values ($\bar{Z}$) estimates the log of the population median.

$$\bar{Z} \text{ estimates } \log[\text{Median}(Y)]$$

$$e^{\bar{Z}} \text{ estimates } \text{Median}(Y)$$

The Natural Logarithm Transformation

4