Concepts: Transformations for ANOVA models

20190930 – Sleuth3 Sections 3.5 and 5.5

Context

- Transformations can sometimes help with the following issues:
  - non-normal distributions within each group (but skewness is only a problem if it is very serious)
  - lack of equal variance for all groups
  - outliers (but usually only if this is a side effect of serious skewness)
- The most common transformations (that we'll consider in this class) work for positive numbers only.

The Ladder of Powers

- Imagine a "ladder of powers" of y (or x): We start at y and go up or down the ladder.

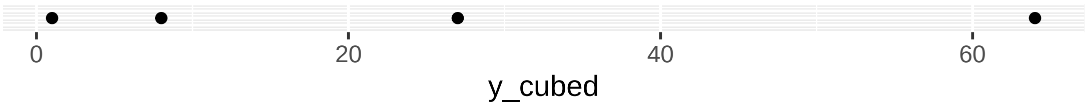
Transformation	R Code	Comments
$\vdots$		
e^y	`exp(y)`	Exactly where on the ladder the exponential transformation belongs depends on the magnitude of the data, but somewhere around here...
y^2	`y^2`	
y		Start here (no transformation)
$\sqrt{y}$	`sqrt(y)`	
$y^{\text{“}0\text{”}}$	`log(y)`	We use $\log(y)$ here
$-1/\sqrt{y}$	`-1/sqrt(y)`	The $-$ keeps the values of y in order
$-1/y$	`-1/y`	
$-1/y^2$	`-1/y^2`	
$\vdots$		

Some (minimal) facts about logarithms and exponentials

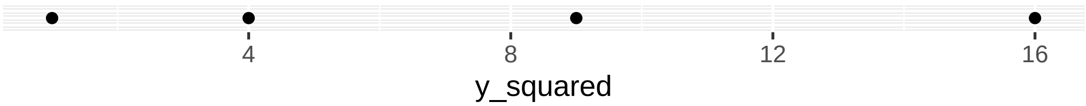
- Foundations:
  - In this class the base of our logarithms is e
  - Notation: $\exp(x) = e^x$
- log() and exp() are inverses
  - $\log(\exp(x)) = x$
  - $\exp(\log(x)) = x$
- They are useful because they convert multiplication to addition, and addition to multiplication
  - $\log(a \cdot b) = \log(a) + \log(b)$
  - $\exp(a + b) = \exp(a) \cdot \exp(b)$

- Which direction?
  - If a variable is skewed right, move it down the ladder (pull down large values)
  - If a variable is skewed left, move it up the ladder (pull up small values)

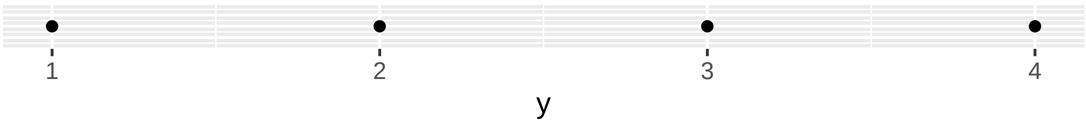
2 Steps Up from Goal: y^3 is very skewed right



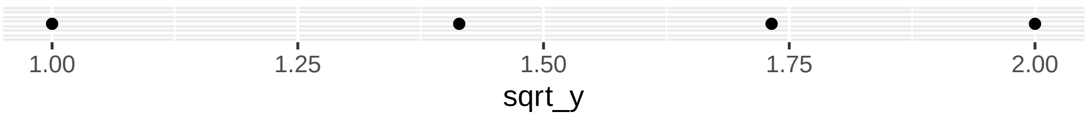
1 Step Up from Goal: y^2 is slightly skewed right



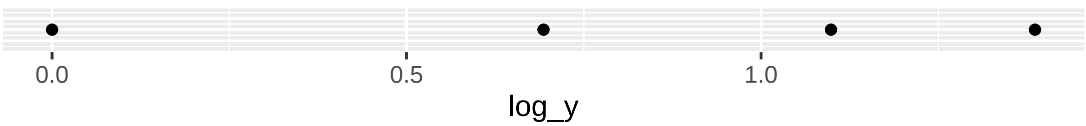
Goal: y is symmetric



1 Step Down from Goal: $\sqrt{y}$ is slightly skewed left

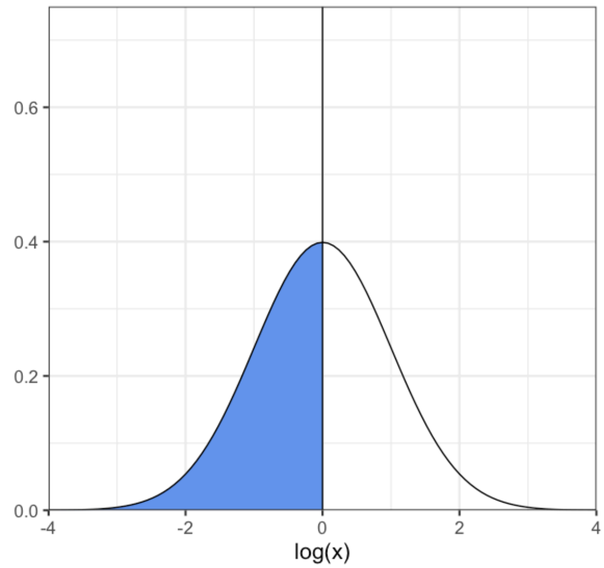


2 Steps Down from Goal: log(y) is very skewed left



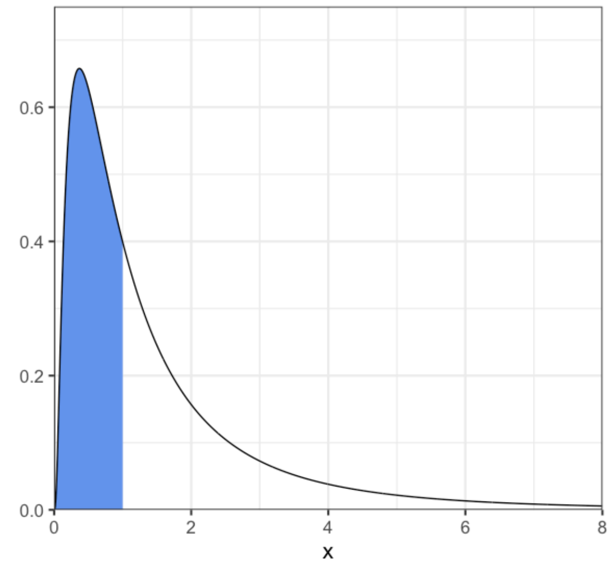
Interpretation of a single mean on the original scale
(valid for all transformations, illustrated with log here)

Shaded area is 0.5 in both cases.



On transformed scale, median = mean (ideally, distribution is close to symmetric after transformation)

Example: on log scale, median is 0



Median on original scale is exponential transformation of median on log scale.

Example: on original scale, median is $e^0 = 1$

Interpretation of a difference between means on the original scale
(valid for log transformation only!)

$$
\begin{aligned}
&\exp\{\textbf{Mean Group 2 on log scale} - \textbf{Mean Group 1 on log scale}\}\\
&\quad = \exp\{\log(\textbf{Median group 2}) - \log(\textbf{Median group 1})\}\\
&\quad = \exp\left\{\log\left(\frac{\textbf{Median group 2}}{\textbf{Median group 1}}\right)\right\}\\
&\quad = \frac{\textbf{Median group 2}}{\textbf{Median group 1}}
\end{aligned}
$$

Rearranging, we obtain:

$$\textbf{Median group 2} = \textbf{Median group 1} \times \exp(\textbf{Mean Group 2 on log scale} - \textbf{Mean Group 1 on log scale})$$

Equivalently, …

$$\textbf{Median group 1} = \textbf{Median group 2} \times \exp(\textbf{Mean Group 1 on log scale} - \textbf{Mean Group 2 on log scale})$$