Interpreting Log Transformations

David Gerard

2018-12-07

Objectives

- Interpret Log-transformations of either the explanatory of response variable.
- Ch 8 in the book.

No log on X, no log on Y

No log on X, no log on Y: Interpretation

- Model: $\mu(Y|X) = \beta_0 + \beta_1 X$

No log on X, no log on Y: Interpretation

- Model: $\mu(Y|X) = \beta_0 + \beta_1 X$
- So at $X + 1$ the mean is

$$\begin{aligned}\mu(Y|X+1) &= \beta_0 + \beta_1(X+1) \\ &= \beta_0 + \beta_1 X + \beta_1 \\ &= \mu(Y|X) + \beta_1\end{aligned}$$

No log on X, no log on Y: Interpretation

- Observational study interpretation: β_1 is the mean difference in the Y's when the X's are only one unit apart.
- Randomized experiment interptation: β_1 is the increase in Y when X is increased by one unit.

No log on X, log on Y

No log on X, log on Y: Interpretation

- Model: $\mu(\log(Y)|X) = \beta_0 + \beta_1 X$

No log on X, log on Y: Interpretation

- Model: $\mu(\log(Y)|X) = \beta_0 + \beta_1 X$
- If *residuals* are fairly symmetric, this means

$$\begin{aligned} Median(\log(Y)|X) &= \beta_0 + \beta_1 X \\ \Rightarrow Median(Y|X) &= \exp(\beta_0 + \beta_1 X) \\ &= \exp(\beta_0)\exp(\beta_1 X) \end{aligned}$$

No log on X, log on Y: Interpretation

- Model: $\mu(\log(Y)|X) = \beta_0 + \beta_1 X$
- If *residuals* are fairly symmetric, this means

$$
\begin{aligned}
Median(\log(Y)|X) &= \beta_0 + \beta_1 X \\
\Rightarrow Median(Y|X) &= \exp(\beta_0 + \beta_1 X) \\
&= \exp(\beta_0)\exp(\beta_1 X)
\end{aligned}
$$

- So at $X + 1$ we have

$$
\begin{aligned}
Median(Y|X+1) &= \exp(\beta_0)\exp(\beta_1(X+1)) \\
&= \exp(\beta_0)\exp(\beta_1 X + \beta_1) \\
&= \exp(\beta_0)\exp(\beta_1 X)\exp(\beta_1) \\
&= Median(Y|X)\exp(\beta_1)
\end{aligned}
$$

No log on X, log on Y: Interpretation

- Observational study interpretation: $\exp(\beta_1)$ is the ratio of medians of the Y's when they differ only by one unit of the X's.
- Randomized experiment interptation: $\exp(\beta_1)$ is the multiplicative change in Y when X is increased by one unit.

No log on X, log on Y: Example

- Voltage and Breakdown: $\hat{\beta}_1 = -0.51$, 95% CI (-0.62, -0.39).
- “Increasing voltage of 1 kV results in a multiplicative change of $\exp(-0.51) = 0.6$”"
- “Breakdown time at 28 kV is 60% that of 27 kV”
- 95% confidence interval of the multiplicative effect is $(\exp(-0.62), \exp(-0.39)) = (0.54, 0.68)$.

No log on X, log on Y: Example

$$
\begin{aligned}
&Median(Y|X+1) - Median(Y|X) \\
&= Median(Y|X)\exp(\beta_1) - Median(Y|X) \\
&= (\exp(\beta_1) - 1)Median(Y|X) \\
&= (0.6 - 1)Median(Y|X) \\
&= -0.4Median(Y|X)
\end{aligned}
$$

- "Increasing kV by 1 decreases median breakdown time by 40%."

log on X, no log on Y

log on X, no log on Y: Interpretation

- Model $\mu(Y|X) = \beta_0 + \beta_1 \log(X)$

log on X, no log on Y: Interpretation

- Model $\mu(Y|X) = \beta_0 + \beta_1 \log(X)$
- Observe what happens to X when we multiply it by 2.

$$\begin{aligned}
\mu(Y|2X) &= \beta_0 + \beta_1 \log(2X) \\
&= \beta_0 + \beta_1(\log(X) + \log(2)) \\
&= \beta_0 + \beta_1 \log(X) + \beta_1 \log(2) \\
&= \mu(Y|X) + \beta_1 \log(2)
\end{aligned}$$

log on X, no log on Y: Interpretation

- Observational study interpretation: $\beta_1 \log(2)$ is the mean difference in the Y's when the ratio of the X's is 2.
- Randomized experiment interptation: $\beta_1 \log(2)$ is the increase in Y when X is doubled.
- Similar interpretation for $\beta_1 \log(10)$ and a 10-fold increase in X.

log on X, no log on Y: Example

- pH and time in meat after slaughter
- $Y_i = \beta_0 + \beta_1 \log(X_i)$
- $\hat{\beta}_1 = -0.726$, 95% CI (-0.805, -0.646).
- If we double the time, then the *pH* decreases by $0.726 \log(2) = 0.503$
- 95% CI: $(-0.805 \log(2), -0.646 \log(2)) = (-0.558, -0.448)$.

log on X, log on Y

log on X, log on Y: Interpretation

- Combination of the past two interpretations

log on X, log on Y: Interpretation

- Combination of the past two interpretations
- Simplifying the model:

$$\begin{aligned} Median(Y|X) &= \exp(\beta_0)\exp(\beta_1 \log(X)) \\ &= \exp(\beta_0)\exp(\log(X^{\beta_1})) \\ &= \exp(\beta_0)X^{\beta_1} \end{aligned}$$

log on X, log on Y: Interpretation

- Combination of the past two interpretations
- Simplifying the model:

$$
\begin{aligned}
\mathit{Median}(Y|X) &= \exp(\beta_0)\exp(\beta_1 \log(X)) \\
&= \exp(\beta_0)\exp(\log(X^{\beta_1})) \\
&= \exp(\beta_0)X^{\beta_1}
\end{aligned}
$$

- See how $\mathit{Median}(Y|X)$ when you double X

$$
\begin{aligned}
\mathit{Median}(Y|2X) &= \exp(\beta_0)(2X)^{\beta_1} \\
&= \exp(\beta_0)X^{\beta_1}2^{\beta_1} \\
&= \mathit{Median}(Y|X)2^{\beta_1}
\end{aligned}
$$

log on X, log on Y: Interpretation

- Observational study interpretation: 2^{β_1} is the ratio of medians of the Y's when the ratio of the X's is 2.
- Randomized experiment interptation: 2^{β_1} is the multiplicative change in Y when X is doubled.
- Similar interpretation for 10^{β_1} and a 10-fold increase in X.

Other Transformations

- Might make residuals look better but don't have nice interpretations.
- If goal is
  1. Prediction
  2. Just answering **if** there is an association and you don't care what it is.

  Then you can try other transformations (because interpretation does not matter).
- $1/Y$ tends to fix more extreme non-constant variance.
- $\sqrt{Y}$ tends to fix less extreme non-constant variance.