

INTERPRETING COEFFICIENTS IN MODELS WITH $\log(\cdot)$ TERMS

MAX H. FARRELL
BUS 41100
DECEMBER 4, 2017

This handout explains how to interpret regression coefficients when one or more variable is log transformed. We will first look at linear regression and then logistic regression. We will only look at simple linear/logistic regression for simplicity. Everything carries over unchanged to multiple regression, with the usual addition that interpretations are holding everything else fixed. In a simple regression we have an outcome Y and one covariate X , and we can take the log of either or both. Thus, we have four possibilities: no transformation, only X , only Y , or both.

0 Preliminaries

The symbol Δ means “change in”. Our goal in regression is to find the general trend in how Y changes as X changes, that is $\Delta Y/\Delta X$. In calculus, this is the derivative of Y with respect to X .

Two key properties of the log function are

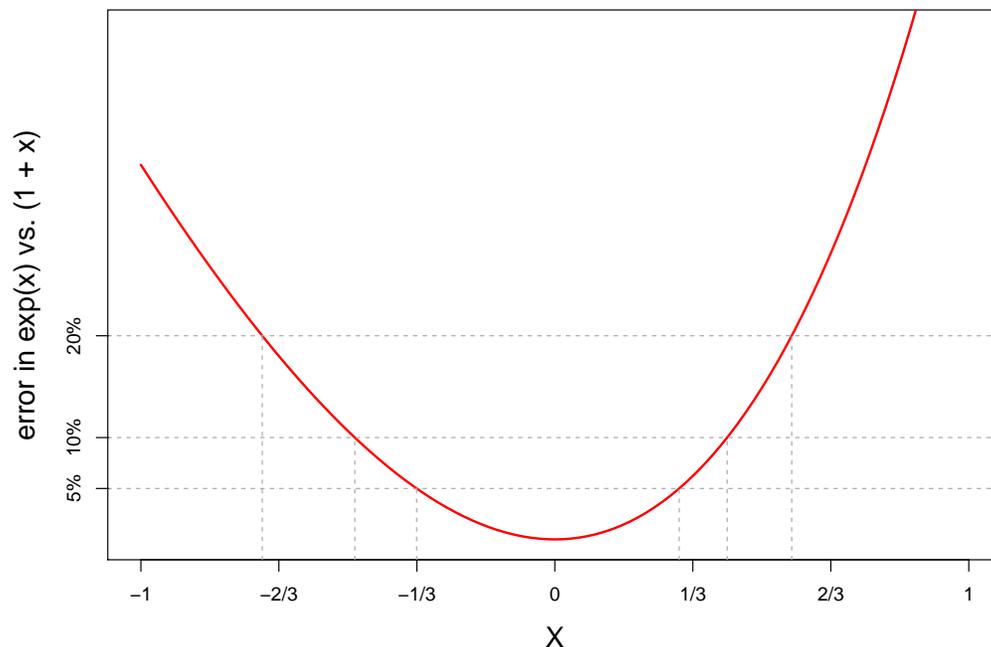
$$\log(x) - \log(y) = \log(x/y) \quad (1)$$

$$\log(1 + x) \approx x \text{ if } x \text{ is “small”} \quad (2)$$

The second property can also be stated as

$$e^x \approx 1 + x \text{ if } x \text{ is “small”} \quad (3)$$

The quality of this approximation gets worse rapidly when x is “large”. See the graph in Figure 1 below; you can judge for yourself if you think it is appropriate to use in a particular set of results.



1 Linear Regression

1.1 level-level model: neither Y nor X is transformed

The model is $Y = \beta_0 + \beta_1 X + \varepsilon$. If X changes from X to $X + \Delta X$, so that the change in X is ΔX , what is the change in Y ?

$$\Delta Y = (\beta_0 + \beta_1(X + \Delta X) + \varepsilon) - (\beta_0 + \beta_1(X) + \varepsilon) = \beta_1 \Delta X \quad \Rightarrow \beta_1 = \frac{\Delta Y}{\Delta X}$$

So for a one-unit change in X , Y changes by β_1 units (in expectation!).

1.2 level-log model: only X is transformed

The model is $Y = \beta_0 + \beta_1 \log(X) + \varepsilon$. If X changes:

$$\begin{aligned} \Delta Y &= (\beta_0 + \beta_1 \log(X + \Delta X) + \varepsilon) - (\beta_0 + \beta_1 \log(X) + \varepsilon) = \beta_1 (\log(X + \Delta X) - \log(X)) \\ &= \beta_1 \log\left(\frac{X + \Delta X}{X}\right) = \beta_1 \log\left(1 + \frac{\Delta X}{X}\right) \approx \beta_1 \left(\frac{\Delta X}{X}\right) \quad \Rightarrow \beta_1 \approx \frac{\Delta Y}{(\Delta X/X)} \end{aligned}$$

So β_1 is the expected change in Y for a one-unit increase in the quantity $(\Delta X/X)$, which is the percent change in X . That is, β_1 is the expected change in Y for a 100% increase in X .

Note, people usually report the change in Y for a 1% change in X , so you have to **divide** your estimate of β_1 by 100.

1.3 log-level model: only Y is transformed

The model is $\log(Y) = \beta_0 + \beta_1 X + \varepsilon$. Transforming this we have $Y = e^{\beta_0 + \beta_1 X + \varepsilon}$, and we use this form to find ΔY when X changes, by plugging it in in two places:

$$\Delta Y = e^{\beta_0 + \beta_1(X + \Delta X) + \varepsilon} - e^{\beta_0 + \beta_1 X + \varepsilon} = e^{\beta_1(\Delta X)} Y - Y \quad \Rightarrow \quad \frac{Y + \Delta Y}{Y} = e^{\beta_1(\Delta X)}.$$

Taking the log of both sides: $\log(e^{\beta_1(\Delta X)}) = \beta_1(\Delta X)$ and

$$\log\left(\frac{Y + \Delta Y}{Y}\right) = \log\left(1 + \frac{\Delta Y}{Y}\right) \approx \frac{\Delta Y}{Y} \quad \Rightarrow \beta_1 \approx \frac{\Delta Y}{\Delta X} \frac{1}{Y}$$

So β_1 is the change in $(\Delta Y/Y)$ for a one-unit change in X . To convert this to a percent, we have to **multiply** by 100. So if $b_1 = 0.3$, then a one-unit change in X means we expect a 30% change in Y .

1.4 log-log model: both Y and X are transformed

The model is $\log(Y) = \beta_0 + \beta_1 \log(X) + \varepsilon$. Transforming this we have $Y = e^{\beta_0 + \beta_1 \log(X) + \varepsilon}$. Using this, and the algebra steps from §1.2 and §1.3, we find

$$\Rightarrow \beta_1 \approx \frac{\Delta Y/Y}{\Delta X/X} = \frac{\Delta Y}{\Delta X} \frac{X}{Y}$$

which is the **elasticity** of Y with respect to X : for a 1% change in X we expect a $\beta_1\%$ change in Y .

2 Logistic Regression

The logistic regression model said that the log odds ratio was a linear function:

$$\log \left(\frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]} \right) = \beta_0 + \beta_1 X$$

The “ Y ” side is always log transformed, so we just have to interpret β_1 when X is log transformed or not.

To simplify the notation, we will write

$$OR(X) = \frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]}$$

Everything will be in terms of odds ratios here, but remember, if you want the actual probabilities you can always solve $OR = p/(1 - p)$, where $p = \mathbb{P}[Y = 1|X]$.

2.1 X is not transformed

We did this in class. Change X to $X + \Delta X$, so that the change in X is ΔX . The logistic model implies:

$$\log \left(\frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]} \right) = \beta_0 + \beta_1 X \quad \Leftrightarrow \quad OR(X) := \frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]} = \exp \{ \beta_0 + \beta_1 X \}.$$

We divide the odds ratios to get

$$\frac{OR(X + \Delta X)}{OR(X)} = \frac{\exp \{ \beta_0 + \beta_1 (X + \Delta X) \}}{\exp \{ \beta_0 + \beta_1 X \}} = e^{\beta_1 \times \Delta X}$$

Therefore for a ΔX unit change in X , we multiply the odds ratio by $e^{\beta_1 \times \Delta X}$:

$$OR(X + \Delta X) = OR(X) e^{\beta_1 \times \Delta X}.$$

In particular, in class, we set $\Delta X = 1$, and found that you multiply the odds ratio by e^{β_1} . Thus we found that exponentiating the coefficients aided in interpretation

2.2 X is log transformed

The logistic model says in this case

$$\log\left(\frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]}\right) = \beta_0 + \beta_1 \log(X).$$

This looks just like the log-log model from Section 1.4 above (and week 4 in class), and therefore β_1 can be interpreted as the percent change in “ Y ” for a 1% change in X , except here the role of “ Y ” is played by the odds ratio.

So when you estimate a logistic regression with $\log(X)$, the coefficient estimate b_1 does not need to be exponentiated to be interpreted: b_1 is the **percent** change in the odds ratio for a 1% increase in X . But note carefully that this is the **percent** change in the odds ratio, whereas in class, and in Section 2.1 immediately above, we were talking about just the change in the odds ratio. These are different

$$\text{percent change} = \frac{\Delta OR}{OR} \quad \frac{OR(X + \Delta X)}{OR(X)} = \frac{OR(X) + \Delta OR}{OR(X)} = 1 + \frac{\Delta OR}{OR}.$$

To link the two, consider a one percent change in X , i.e. changing X to $X + 0.01X$, and re-do the logic from above/class. The logistic model implies:

$$\log\left(\frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]}\right) = \beta_0 + \beta_1 \log(X) \quad \Leftrightarrow \quad OR(X) := \frac{\mathbb{P}[Y = 1|X]}{\mathbb{P}[Y = 0|X]} = \exp\{\beta_0 + \beta_1 \log(X)\}.$$

Dividing these, but comparing the odds ratio at X and $1.01 \times X$, we get

$$\begin{aligned} \frac{OR(1.01X)}{OR(X)} &= \frac{\exp\{\beta_0 + \beta_1 \log(1.01X)\}}{\exp\{\beta_0 + \beta_1 \log(X)\}} = \exp\{\beta_1 (\log(1.01X) - \log(X))\} \\ &= \exp\left\{\beta_1 \log\left(\frac{1.01X}{X}\right)\right\} = \exp\{\beta_1 \log(1.01)\} \approx \exp\{\beta_1 \times 0.01\} \end{aligned}$$

Therefore to get the new odds ratio, we take the old one and multiply by $\exp\{\beta_1 \times 0.01\}$. This is the analogue of above/class, where for a one unit increase in X we multiplied the odds ratio by $\exp\{\beta_1\}$. This is the **change** in the odds ratio, but it is **not** the percent change.

To get the percent change, use the fact from above that

$$\frac{OR(1.01X)}{OR(X)} = 1 + \frac{\Delta OR}{OR},$$

where the second term is the percent change. Plugging this in, and using the fact that $e^z \approx 1 + z$

$$\frac{OR(1.01X)}{OR(X)} = 1 + \frac{\Delta OR}{OR} = \exp\{\beta_1 \times 0.01\} \approx 1 + \beta_1 \times 0.01 = 1 + \beta_1 \frac{\Delta X}{X},$$

Then cancel the 1 from both sides, and you get

$$\frac{\Delta OR}{OR} = \beta_1 \frac{\Delta X}{X},$$

so that β_1 is the percent change in the OR for a 1% change in X , just like it was before.