A summary of the interpretation of regression coefficients for the untransformed variables and for log transformations.

Simple linear regression

The "casual" argument looks like this:

$$\begin{aligned} y' &= \beta_0 + \beta_1(x+1) \\ y &= \beta_0 + \beta_1 x \\ \Delta y = y' - y &= \beta_1 \cdot 1 \end{aligned}$$

But in fact on the right hand side we are really dealing with expected values, so we should write instead that for the model

$$y = \beta_0 + \beta_1 x + \epsilon, \;\; \epsilon \sim N(0, \sigma^2)$$

we have

$$\begin{aligned} E[y|x+1] &= \beta_0 + \beta_1(x+1) \\ E[y|x] &= \beta_0 + \beta_1 x \\ \Delta E[y] = E[y|x+1] - E[y|x] &= \beta_1 \cdot 1 \end{aligned}$$

so that β_1 is the change in the *expected value* of y, for a one unit change in x.

In casual language we might say β_1 is the change in y for a one-unit change in x, but we know that's not quite true. We don't know the change in y exactly (because of ϵ) but we do know about the expected value of the change.

Generalization to multiple regression

If we focus on the change from x_j to $x_j + 1$ in the regression model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

we have

$$\begin{aligned} E[y|x_j+1] &= \beta_0 + \beta_1 x_1 + \cdots + \beta_j(x_j+1) + \cdots \beta_p x_p \\ E[y|x_j] &= \beta_0 + \beta_1 x_1 + \cdots + \beta_j x_j + \cdots \beta_p x_p \\ \Delta_j E[y] = E[y|x_j+1] - E[y|x_j] &= \beta_j \cdot 1 \end{aligned}$$

so that β_j is the change in the *expected value* of y, for a one unit change in x_j, *holding the other x's fixed*. Since the adjustment is always like this for multiple regression, we will only consider simple regression examples for the rest of these notes.

log(Y)

Now we want to consider the model

$$\log y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

Following what we did for simple linear regression, we want to consider

$$\Delta E[\log y] = E[\log y | x+1] - E[\log y | x].$$

To simplify the notation a bit and prepare for the application of a Taylor approximation, we will write

$$\log(y + \Delta y) = \beta_0 + \beta_1 (x+1) + \epsilon$$

for the value observed at $x + 1$ and

$$\log(y) = \beta_0 + \beta_1 x + \epsilon'$$

for the value observed at x. Taking expected values to get rid of the ϵ's, we get

$$\begin{aligned} E[\log(y + \Delta y)] &= \beta_0 + \beta_1(x+1) \\ E[\log(y)] &= \beta_0 + \beta_1 x \\ \Delta E[\log y] = E[\log(y + \Delta y)] - E[\log(y)] &= \beta_1 \cdot 1 \end{aligned}$$

Continuing the calculation to get a "percent" interpretation, we have

$$\begin{aligned} \beta_1 &= E[\log(y + \Delta y)] - E[\log(y)] \\ &= E\left[\log\left(1 + \frac{\Delta y}{y}\right)\right] \\ &\approx E\left[\frac{\Delta y}{y}\right] \qquad \text{Taylor approx: } \log(1+u) = u - u^2/2 \pm \cdots \end{aligned}$$

So, for the model

$$\log y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

we can say β_1 is the *expected value of* the fractional change in y for a 1 unit change in x. It is perhaps more colorful to say that

"$\beta_1 \times 100\%$ is the expected percent change in y for a one-unit change in x."

In casual language we could say "a one-unit change in x is associated with a $\beta_1 \times 100\%$ change in y, but we now know that's not quite accurate; we don't really know what the change in y will be (because of ϵ) but we do know about the expected value of the (percent) change.

log(X)

Now let's consider the model

$$y = \beta_0 + \beta_1 \log x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

We will consider a small change Δx in x:

$$\begin{aligned} E[y|x + \Delta x] &= \beta_0 + \beta_1 \log(x + \Delta x) \\ E[y|x] &= \beta_0 + \beta_1 \log x \\ \Delta E[y] = E[y|x + \Delta x] - E[y|x] &= \beta_1 \cdot \log\left(1 + \frac{\Delta x}{x}\right) \\ &\approx \beta_1 \cdot \left(\frac{\Delta x}{x}\right) \qquad \text{Taylor approx: } \log(1+u) = u - u^2/2 \pm \cdots \end{aligned}$$

If we put $\Delta x = 0.01x$ (a one-percent change in x), we get

$$\Delta E[y] = E[y|(1.01)x] - E[y|x] = \beta_1 \cdot (0.01)$$

Thus we can say that *"$\beta_1 \times (0.01)$ is the change in the expected value of y for a 1% change in x."*

Both log(X) and log(Y)

Next we consider the model

$$\log y = \beta_0 + \beta_1 \log x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

Combining the calculations for $\log y$ and $\log x$ we get

$$\begin{aligned} E[\log(y + \Delta y)] &= \beta_0 + \beta_1 \log(x + \Delta x) \\ E[\log(y)] &= \beta_0 + \beta_1 \log(x) \\ \Delta E[\log y] = E[\log(y + \Delta y)] - E[\log(y)] &\approx \beta_1 \cdot \left(\frac{\Delta x}{x}\right) \end{aligned}$$

so

$$\beta_1 \cdot \left(\frac{\Delta x}{x}\right) = E[\log(y + \Delta y)] - E[\log(y)] = E\left[\log\left(1 + \frac{\Delta y}{y}\right)\right] \approx E\left[\frac{\Delta y}{y}\right]$$

Putting $\Delta x = 0.01x$, we get that $\beta_1 \cdot (0.01)$ is the expected fractional change in y for a 1% change in x. Multiplying through by 100% we get that $\beta_1 \cdot (0.01) \cdot 100\%$ is the expected percent change in y for 1% change in x. Since (0.01)(100) = 1, we can say more succinctly that

"β_1 (itself!) is the expected percent change in y for a 1% change in x."

So this at least is nice and simple!

Logistic regression

Untransformed x: For logistic regression with an untransformed x, we have

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x$$

where $p = P[y = 1|x]$. In order to keep things clear, let's write $p(x) = P[y = 1|x]$. Following what we did for simple linear regression we get

$$\begin{aligned}
\text{logit}\, p(x+1) = \log \frac{p(x+1)}{1-p(x+1)} &= \beta_0 + \beta_1 (x+1) \\
\text{logit}\, p(x) = \log \frac{p(x)}{1-p(x)} &= \beta_0 + \beta_1 x \\
\Delta \text{logit}\, p = \text{logit}\, p(x+1) - \text{logit}\, p(x) &= \beta_1 \cdot 1
\end{aligned}$$

so that β_1 *is the (additive) change in logit p, for a one unit change in x.*

Exponentiating both sides and simplifying, we get

$$\frac{p(x+1)}{1-p(x+1)} = \frac{p(x)}{1-p(x)} \cdot e^{\beta_1}$$

so that e^{β_1} *is the (multiplicative) change in the odds of* $y = 1$*, for a one unit change in x.*

Transforming to log(x): Finally let's consider the model

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 \log x$$

Following what we did for $\log x$ in simple regression we get

$$\begin{aligned}
\text{logit}\, p(x+\Delta x) = \log \frac{p(x+\Delta x)}{1-p(x+\Delta x)} &= \beta_0 + \beta_1 \log(x+\Delta x) \\
\text{logit}\, p(x) = \log \frac{p(x)}{1-p(x)} &= \beta_0 + \beta_1 \log(x) \\
\Delta \text{logit}\, p &= \beta_1 \log\left(1 + \frac{\Delta x}{x}\right) \approx \beta_1 \cdot \left(\frac{\Delta x}{x}\right)
\end{aligned}$$

Putting $\Delta x = 0.01x$ again, we get that $\beta_1 \cdot (0.01)$ *is the (additive) change in logit p, for a 1% change in x.*

Exponentiating again, we get that $e^{\beta_1 \cdot (0.01)}$ *is the (multiplicative) change in the odds of* $y = 1$*, for a 1% change in x.*

Following the logic of the case of simple regression where both x and y are replaced with logaritthms, we could also say that β_1 *is the expected percent change in the odds of* $y = 1$*, for a 1% change in x*. If you use the Taylor approximation $e^u = 1 + u + \cdots$ with $u = \beta_1 \cdot (0.01)$, you can also get this interpretation from the multiplicative interpretation in the previous paragraph.