Interpreting Regression Coefficients for Log-Transformed Variables

Statnews #83

Cornell Statistical Consulting Unit

Created June 2012. Last updated September 2020

Introduction

Log transformations are one of the most commonly used transformations, but interpreting results of an analysis with log-transformed data may be challenging. This newsletter focuses on how to obtain estimated parameters of interest and how to interpret the coefficients in a regression model involving log-transformed variables. A log transformation is often useful for data which exhibit right skewness (positively skewed), and for data where the variability of residuals increases for larger values of the dependent variable. When some variables are log-transformed, estimating parameters of interest based on the model may involve more calculation than simply taking the anti-log of certain regression coefficients.

The log-normal distribution

To properly back transform into the original scale we need to understand some details about the log-normal distribution. In probability theory, a log-normal distribution is the distribution of the random variable Y when $\ln(Y)$ follows a normal distribution with mean μ and variance σ^2. If we think of Y as the response variable in a regression model, then log-transforming the response variable and fitting a linear regression is equivalent to assuming that $\ln(Y)$ follows a normal distribution. So it will be helpful to understand the behavior of Y in terms of the parameters of the normally distributed variable $\ln(Y)$.

If $\ln(Y)$ is normally distributed with mean μ and variance σ^2, then the following statements are true:

- The mean of Y is $e^{\mu+\sigma^2/2}$
- The median of Y is e^{μ}
- The variance of Y is $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$

Suppose we fit a linear regression model with predictors $x_1, \ldots, x_p$ and log-transformed response variable $\ln(Y)$. With typical modeling assumptions this means that $\ln(Y)$ has a normal distribution with mean $\mu = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ and variance σ^2. Given the coefficient

estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$, the predicted value for the mean of $\ln(Y)$ is $\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$. It is important to note that exponentiating this predicted value does not provide an estimate of the mean of Y. Given the three facts stated above, an estimate of the mean of Y is given by

$$e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p + \hat{\sigma}^2/2},$$

where $\hat{\sigma}^2$ is the residual mean squared error from the fitted regression model.

Coefficient interpretation

Interpreting parameter estimates in a linear regression when some variables are log-transformed is not always straightforward. The standard interpretation of a regression parameter β_j is that a one-unit change in the corresponding predictor x_j is associated with β_j units of change in the expected value of the response variable, holding all other predictors constant.

The interpretation of regression coefficients when one or more variables are log-transformed depends on whether the dependent variable, independent variable, or both are transformed. To understand each of these cases, consider an example in which weight is the dependent variable and height is the only independent variable.

Only the dependent variable is transformed

Linear change in the independent variable is associated with multiplicative change in the dependent variable.

Suppose the fitted model is

$$\ln(\text{weight}) = 2.14 + 0.00055\text{height}$$

The estimated coefficient for height is $\hat{\beta}_1 = 0.00055$, so we would say that an increase of one unit in height is associated with a $100 \times (e^{\hat{\beta}_1} - 1) \approx 0.055$ percent change in weight.

Explanation

Given the model $\ln(y) = \beta_0 + \beta_1 x$, consider increasing x by one unit. If we call y_{new} the value of y after increasing x by one unit, then $\ln(y_{\text{new}}) = \beta_0 + \beta_1(x + 1) = \ln(y) + \beta_1$. Therefore $\ln(y_{\text{new}}) - \ln(y) = \beta_1$, or $e_1^{\beta} = y_{\text{new}}/y$, and

$$100 \times \left(\frac{y_{\text{new}}}{y} - 1\right) = 100 \times \left(\frac{y_{\text{new}} - y}{y}\right) = 100 \times (e^{\beta_1} - 1)$$

is the percent change in y associated with a one-unit increase in x.

Only the independent variable is transformed

Multiplicative change in the independent variable is associated with linear change in the dependent variable.

Fitted model:

$$\text{weight} = 3.94 + 1.16\ln(\text{height})$$

Here $\hat{\beta}_1 = 1.16$ and we would say that a one-percent increase in height is associated with an increase of $\hat{\beta}_1 \ln(1.01) \approx 0.0115$ in weight.

Explanation

The model is $y = \beta_0 + \beta_1 \ln(x)$ and we consider increasing x by one percent, i.e. $x_{\text{new}} = 1.01x$. Then

$$y_{\text{new}} = \beta_0 + \beta_1 \ln(x_{\text{new}}) = \beta_0 + \beta_1 \ln(1.01x) = \beta_0 + \beta_1 \ln(x) + \beta_1 \ln(1.01) = y + \beta_1 \ln(1.01).$$

This means that $y_{\text{new}} - y = \beta_1 \ln(1.01)$, so the increase in y associated with a one-percent increase in x is $\beta_1 \ln(1.01.)$

Both the independent and dependent variable are transformed

Multiplicative change in the independent variable is associated with multiplicative change in the dependent variable.

Fitted model:

$$\ln(\text{weight}) = 1.69 + 0.11 \ln(\text{height})$$

In this case, $\hat{\beta}_1 = 0.11$ and we would say that a one-percent increase in height is associated with a $100 \times (1.01^{\hat{\beta}_1} - 1)$ percent change in weight, or about a 0.11 percent change in weight.

Explanation

Now the model is $\ln(y) = \beta_0 + \beta_1 \ln(x)$. Consider increasing x by one percent, i.e. $x_{\text{new}} = 1.01x$. Then

$$\ln(y_{\text{new}}) = \beta_0 + \beta_1 \ln(1.01x) = \beta_0 + \beta_1 \ln(x) + \beta_1 \ln(1.01) = \ln(y) + \beta_1 \ln(1.01).$$

Therefore $\ln(y_{\text{new}}) - \ln(y) = \beta_1 \ln(1.01)$ and

$$e^{\ln(y_{\text{new}}) - \ln(y)} = \frac{y_{\text{new}}}{y} = e^{\beta_1 \ln(1.01)} = 1.01^{\beta_1}$$

so that the percent change in y associated with a one-percent increase in x is

$$100 \times \left(\frac{y_{\text{new}} - y}{y} \right) = 100 \times (1.01^{\beta_1} - 1)$$

As always if you would like assistance with this topic or any other statistical consulting question, feel free to contact statistical consultants at CSCU.

Author: Jing Yang