Worksheet: log transformations - Solution

Stat 230 - chapter 8 - St. Clair

Let $b > 0$ and $x > 0$. The logarithm (base-b) of x is denoted $\log_b(x)$ and equal to

$$\log_b(x) = a$$

where a tells us what power we must raise b to to obtain the value x:

$$b^a = x$$

Easy examples are: $b = 2$, $x = 8$ and $a = 3$,

$$\log_2(8) = 3$$

since $2^3 = 8$. Or using base $b = 10$, then

$$\log_{10}(0.01) = -2$$

since $10^{-2} = 0.01$.

Some basic facts logarithm facts are

$$\log_b(b) = 1$$

since $b^1 = b$ and

$$\log_b(1) = 0$$

since $b^0 = 1$.

Interpreting logged variables

Multiplicative changes in x result in **additive** changes in $\log_b(x)$. If $m > 0$, then

$$\log_b(mx) = \log_b(m) + \log_b(x)$$

For example,

$$\log_2(16) = \log_2(2 \times 8) = \log_2(2) + \log_2(8) = 1 + 3 = 4$$

Inverse (i.e. reversing the log, getting rid of the log, . . .)

The logarithm and exponential functions are inverses of one another. This means we can "get rid" of the log by calculating b raised to the logged-function:

$$b^{\log_b(x)} = x$$

This will be useful in regression when we have a linear relationship between logged-response y and a set of predictors. We need to

For example, suppose we know that

$$\log_2(y) = 3 + 5x$$

To return this to an expression on the original (unlogged) scale of y, we need take both sides raised to the base 2:

$$2^{\log_2(y)} = 2^{3+5x}$$

Simplifying both sides gives

$$y = 2^3 \times 2^{5x}$$

Questions

1. Write the following as the sum of two logarithms. Simplify as much as possible:
   a. $\log_2(2x)$

   answer:
$$\log_2(2x) = \log_2(2) + \log_2(x) = 1 + \log_2(x)$$

   b. $\log_2(0.5x)$

   answer:
$$\log_2(0.5x) = \log_2(0.5) + \log_2(x) = -1 + \log_2(x)$$

   c. $\ln(2x)$ where ln is the natural log (base-e)

   answer:
$$\ln(2x) = \ln(2) + \ln(x) \approx 0.693 + \ln(x)$$

2. Write the following expressions in terms of y, not $\log(y)$. Simplify as much as possible:
   a. $\log_2(y) = 1 - 3x$

   answer:
$$y = 2^{1-3x} = 2^1 \times 2^{-3x} = 2 \times (2^{-3})^x = 2 \times (\frac{1}{8})^x$$

   b. $\log_{10}(y) = -2 + 0.4x$

   answer:
$$y = 10^{-2+0.4x} = 10^{-2} \times 10^{0.4x} = 0.01 \times (10^{0.4})^x \approx 0.01 \times 2.512^x$$

   c. $\ln(y) = 1 - 3x$

   answer:
$$y = e^{1-3x} = e^1 \times e^{-3x} = e \times (e^{-3})^x \approx 2.718 \times 0.050^x$$

3. Write the following expressions in terms of y and x, not $\log(y)$ and $\log(x)$. Simplify as much as possible:
   a. $\log_2(y) = 1 - 3\log_2(x)$

   answer:
$$y = 2^{1-3\log_2(x)} = 2^1 \times 2^{-3\log_2(x)} = 2 \times \left(2^{\log_2(x)}\right)^{-3} = 2 \times x^{-3}$$

   b. $\ln(y) = -2 + 0.4\ln(x)$

   answer:
$$y = e^{-2+0.4\ln(x)} = e^{-2} \times e^{0.4\ln(x)} = e^{-2} \times \left(e^{\ln(x)}\right)^{0.4} \approx 0.135 \times x^{0.4}$$

   c. $\ln(y) = 1 - 3\log_2(x)$

   answer: This example shows that using difference log-bases on x and y make simplification more difficult.
$$y = e^{1-3\log_2(x)} = e^1 \times e^{-3\frac{\ln(x)}{\ln(2)}} = e \times \left[e^{\ln(x)}\right]^{-3/\ln(2)} = e \times x^{-3/\ln(2)} \approx e \times x^{-3.322}$$

   Here we use the change of base property: $\log_2(x) = \dfrac{\ln(x)}{\ln(2)}$

4. Logarithmic model: Regression of Y on $\log(x)$ obtains the following estimated mean of Y:
$$\hat{\mu}(Y \mid x) = 1 - 3\log_2(x)$$

   a. What is the change in estimated mean response if we double the value of x?

answer: The mean when we double x is

$$\hat{\mu}(Y \mid 2x) = 1-3\log_2(2x) = 1-3(\log_2(2)+\log_2(x)) = 1-3(1+\log_2(x)) = 1-3\log_2(x)-3 = \hat{\mu}(Y \mid x)-3$$

The mean response of Y is lowered by 3 units when we double x.

b. What is the change in estimated mean response if we triple the value of x?

answer: The mean when we triple x is

$$\hat{\mu}(Y \mid 3x) = 1-3\log_2(3x) = 1-3(\log_2(3)+\log_2(x)) = 1-3\log_2(x)-3\log_2(3) \approx \hat{\mu}(Y \mid x)-4.755$$

The mean response of Y is lowered by about 4.755 units when we triple x.

c. What is the change in estimated mean response if we reduce the value of x by 20%?

answer: The mean when we reduce x by 20%, i.e. multiply it by 0.80, is

$$\hat{\mu}(Y \mid 0.8x) = 1-3\log_2(0.8x) = 1-3(\log_2(0.8)+\log_2(x)) = 1-3\log_2(x)-3\log_2(0.8) \approx \hat{\mu}(Y \mid x)+0.966$$

The mean response of Y is increased by about 0.966 units when we reduce x by 20%.

5. Exponential model: Regression of $\log_2(Y)$ on x obtains the following estimated *median* of Y:

$$\hat{med}(\log_2(Y) \mid x) = -2+0.4x$$

a. Write the median in terms of Y instead of $\log_2(Y)$. Simplify as much as possible.

answer: Since the median of the logged-Y's equals the log of the median of Y, we simplify like we did in question 2.

$$\hat{med}(Y \mid x) = 2^{\log_2(\hat{med}(Y|x))} = 2^{-2+0.4x} = 2^{-2} \times 2^{0.4x} = 0.25 \times 2^{0.4x}$$

b. What is the multiplicative change in estimated median response if we increase x by 1 unit?

answer: The median when we add 1 to x is

$$\hat{med}(Y \mid x+1) = 0.25\times 2^{0.4(x+1)} = 0.25\times 2^{0.4x+0.4} = 0.25\times 2^{0.4x}\times 2^{0.4} = \hat{med}(Y \mid x)\times 2^{0.4} \approx \hat{med}(Y \mid x)\times 1.320$$

A one unit increase in x results in a 1.32-fold increase in the median of Y.

c. What is the percent change in estimated median response if we increase x by 1 unit?

answer: A multiplicative increase of 1.32 is the same as a 32% increase in the median of Y.

$$100\%\times\frac{\hat{med}(Y \mid x+1)-\hat{med}(Y \mid x)}{\hat{med}(Y \mid x)} = 100\%\times\frac{\hat{med}(Y \mid x)\times 1.320-\hat{med}(Y \mid x)}{\hat{med}(Y \mid x)} = 100\%(1.32-1) = 32\%$$

d. What is the multiplicative change in estimated median response if we decrease x by 2 units?

answer: The median when we subtract 2 from x is

$$\hat{med}(Y \mid x-2) = 0.25\times 2^{0.4(x-2)} = 0.25\times 2^{0.4x}\times 2^{-2(0.4)} = \hat{med}(Y \mid x)\times 2^{-2(0.4)} \approx \hat{med}(Y \mid x)\times 0.574$$

A two unit decrease in x results in a multiplicative decrease of 0.574 in the median of Y.

e. What is the percent change in estimated median response if we decrease x by 2 units?

answer: A multiplicative decrease of 0.574 is the same as a 42.6% decrease in the median of Y.

$$100\%\times\frac{\hat{med}(Y \mid x-2)-\hat{med}(Y \mid x)}{\hat{med}(Y \mid x)} = 100\%\times\frac{\hat{med}(Y \mid x)\times 0.574-\hat{med}(Y \mid x)}{\hat{med}(Y \mid x)} = 100\%(0.574-1) = -42.6\%$$

6. Power model: Regression of $\log_2(Y)$ on $\log_2(x)$ obtains the following estimated *median* of Y:

$$\widehat{median}(\log_2(Y) \mid x) = 1 - 3\log_2(x)$$

a. Write the median in terms of Y and x instead of logs. Simplify as much as possible.

answer: Since the median of the logged-Y's equals the log of the median of Y, we simplify like we did in questions 3 and 5.

$$\hat{med}(Y \mid x) = 2^{1-3\log_2(x)} = 2^1 \times (2^{\log_2(x)})^{-3} = 2 \times x^{-3}$$

b. What is the multiplicative change in estimated median response if we increase x by 50%?

answer: The median of Y when multiple x by 1.5 (50% increase) is:

$$\hat{med}(Y \mid 1.5x) = 2 \times (1.5x)^{-3} = 2 \times x^{-3} \times 1.5^{-3} = \hat{med}(Y \mid x) \times 1.5^{-3} \approx \hat{med}(Y \mid x) \times 0.296$$

Increasing x by 50% results in a 0.296-fold decrease in the median of Y.

c. What is the percent change in estimated median response if we increase x by 50%?

answer: Increasing x by 50% results in a 70.4% decrease in the median of Y

$$100\%(0.296 - 1) = -70.4\%$$

d. What is the multiplicative change in estimated median response if we reduce the value of x by 20%?

answer: The median of Y when multiple x by 0.8 (20% decrease) is:

$$\hat{med}(Y \mid 0.8x) = 2 \times (0.8x)^{-3} = 2 \times x^{-3} \times 0.8^{-3} = \hat{med}(Y \mid x) \times 0.8^{-3} \approx \hat{med}(Y \mid x) \times 1.953$$

Decreasing x by 20% results in a 1.953-fold increase in the median of Y.

e. What is the percent change in estimated median response if we reduce the value of x by 20%?

answer: Decreasing x by 20% results in a 95.3% increase in the median of Y

$$100\%(1.953 - 1) = 95.3\%$$