Day 2, Session 1: Logs/Exponentiation

Brian D. Williamson

EPI/BIOST Bootcamp 2017

25 September 2017

Motivation

Generally, our goal in a statistical analysis is to assess an association between two variables (e.g., smoking and lung cancer). We can do this by investigating whether a summary measure (e.g., mean, median) of our outcome (e.g., lung cancer) is unequal between two groups differing in our predictor of interest (e.g., smoking).

There are two simple ways to tell if two numbers are unequal:

- their difference is not equal to 0, or

- their ratio is not equal to 1.

We choose between these based on a variety of criteria, which fall into two general categories: (1) adequately address the scientific question, and (2) gain desirable statistical properties.

Motivation

Differences:

- are generally easier to understand, and
- are better for describing the scientific importance of many comparisons
  - You probably always want \$1,000,000 more than me, even if I have \$10,000,000 (a ratio of 1.1)

Ratios work well when working with small numbers (disclaimer: these numbers are probably only correct to the order of magnitude, but get the point across); for example:

- In the US, 60–64 year old current or former smokers have a probability of 0.00296 of being diagnosed with lung cancer during the next year
- In the US, 60–64 year old never smokers have a probability of 0.000148 of being diagnosed with lung cancer during the next year
- Difference in incidence rates: 0.002812; ratio of incidence rates: 20!

Motivation

Sometimes, the scientific mechanism dictates that ratios are more generalizeable:

- Interventions or risk factors that affect a rate over time (e.g., HIV incidence)
- Biochemical processes (e.g., rates of absorption, where the rate is proportional to drug concentration)

When ratios are scientifically preferred, we can use the logarithm of the ratio to get back to comparing differences (more on this later).

Common variables

Some variables are almost always log transformed:

- Acidity/alkalinity of an aqueous solution: measured as hydrogen ion concentration, but pH usually reported [$-\log_{10}$(H ion conc.)]

- Concentrations of antibodies or mRNA: these differ by orders of magnitude across people, and within people over time

Properties of exponentiation and logarithms come in handy throughout statistics and data analysis; a solid understanding of the basics goes a long way.

Example:

Exponentiation

Exponentiation corresponds to repeated multiplication, and is:

- the second in the order of operations! (P**E**MDAS)
- composed of two numbers: a base, b, and an exponent, n
  - represented as $b^n = \underbrace{b \times b \times \cdots \times b}_{n \text{ times}}$

Positive exponents multiply the base b a number of times given by n; negative exponents multiply the reciprocal of the base b n times. For example, $2^2 = 2 \times 2$, and $2^{-2} = \left(\frac{1}{2}\right)^2 = \frac{1}{2} \times \frac{1}{2}$.

Properties of exponents

- If we multiply two numbers with the same base, we add their exponents
  - $10^3 \times 10^2 = 10^5$; $10^3 \times 10^{-2} = 10^1$
- If we raise an exponentiated number to a power, we multiply the exponents; $(b^n)^m = b^{n \times m}$
- The exponent can be a fraction (like $\frac{1}{2}$), which gives us the root of the base (if the base is positive)
  - $4^{1/2} = \sqrt{4} = 2$; $81^{1/4} = \sqrt[4]{81} = 3$
- Multiplying different bases: first manipulate the exponent so the bases are equal, then add exponents
  - $2^3 \times 4^5 = 2^3 \times (2^2)^5 = 2^3 \times 2^{10} = 2^{13}$
- For any $b, c \neq 0$: $b^0 = 1$, $(b \times c)^n = b^n \times c^n$

Exponential function

- An important constant: e, approximately 2.718
- Useful as a base for powers
- Define $\exp(x) = e^x$ as the exponential function

Exercise: exponents and the exponential function

1. What is the result of x^2 multiplied by x^3?

2. $(x^{-2})^4 =?$

3. $\exp(x - y) =?$

Solutions: exponents and the exponential function

1. $x^2 \times x^3 = x^5$, since we add the exponents when we multiply
2. $(x^{-2})^4 = x^{-2\times 4} = x^{-8}$
3. $\exp(x - y) = e^{x-y} = e^x \times e^{-y} = e^x/e^y = \exp(x)/\exp(y)$

Logarithms

Logarithms (logs) transform multiplication into addition. This leads to many of their mathematical properties.

Before calculators, to multiply two large numbers (e.g., 1234 and 4747), you would:

1. choose a common base (e.g., 10)
2. convert each number into exponentiated form with this base (e.g., $10^{3.091315}$ and $10^{3.676419}$) using a table of logarithms (usually base 10)
3. add the exponents (e.g., $3.091315 + 3.676419 = 6.767734$)
4. convert back to un-exponentiated form (e.g., $10^{6.767734} = 5857798$)

The logarithm base 10 of a number is just the exponent of the number expressed as a power of 10; the logarithm base 10 of 100 is 2, because $10^2 = 100$.

Logs: definition

More generally, we can define the logarithm base k of a number x, written $\log_k(x)$. If $\log_k(x) = y$, then $k^y = x$.

Common convention in early math courses:

- "log(x)" is $\log_{10}(x)$, and
- "ln(x)" is the *natural log* $\log_e(x)$.

In many scientific applications, "log(x)" is $\log_e(x)$!

- This is also true in most biostatistics courses and software

Some basic properties of logarithms:

- undefined for $x \leq 0$
- increasing: as x increases, $\log_k(x)$ increases
- $\log_k(k) = 1$

Changing bases

Using different bases for logarithms is similar to measuring length in different units (e.g., inches, centimeters). No matter what base you use, $\log(1) = 0$.

This implies that we can convert between bases! This is often useful in science: if you transform a variable using log base e, you can change the base to 10 for a (potentially) more interpretable answer.

We can find the base k logarithm of any number using the most common bases (e and 10): $\log_k(x) = \frac{\log_e(x)}{\log_e(k)} = \frac{\log_{10}(x)}{\log_{10}(k)}$.

Logs: identities

- Multiplication: $\log_b(xy) = \log_b(x) + \log_b(y)$
- Division: for $y \neq 0$, $\log_b(x/y) = \log_b(x) - \log_b(y)$
- Powers: $\log_b(x^p) = p \log_b(x)$
- Roots: for $p \neq 0$, $\log_b(x^{1/p}) = \log_b(x)/p$
- Inverse function: $\log_b(b^x) = x \log_b(b) = x$

Log world

$\exp(\cdot)$ and $\log(\cdot)$

- Recall $\exp(x) = e^x$
- Natural log: $\log(x) = \log_e(x)$
- So $x = \log[\exp(x)]$! And $x = \exp[\log(x)]$!

Exercise: logarithms

1. $\log(xy) = ?$

2. $\log(x/y) = ?$

3. $\log[\exp(2x)] = ?$

4. $\exp[\log(x^2)] = ?$

Solutions: logarithms

1. $\log(xy) = \log(x) + \log(y)$
2. $\log(x/y) = \log(x) - \log(y)$
3. $\log[\exp(2x)] = 2x$
4. $\exp[\log(x^2)] = \exp[2\log(x)] = \exp(2)\exp[\log(x)] = x\exp(2)$

Summary

- Exponentiation: can create terms of higher order (larger exponent) than linear terms (exponent 1)
- Logarithms: turn multiplication into addition, using a base
- Most common base: e
- Useful for transforming data or different types of regression (logistic, Poisson)