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).

Examples

Some variables are almost always log transformed:

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

- 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.

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
    - $4^{1/2} = \sqrt{4} = 2$; $81^{1/4} = \sqrt[4]{81} = 3$
- Multiplying different bases: first manipulate the exponent so the bases are equal, then add exponents
    - $2^3 \times 4^5 = 2^3 \times (2^2)^5 = 2^3 \times 2^{10} = 2^{13}$

Exponent identities

- For all $b, c \neq 0$:
  - $b^{m+n} = b^m \times b^n$
  - $b^{m \times n} = (b^m)^n$
  - $(b \times c)^n = b^n \times c^n$

Example: integer exponent properties and identities

- Take $b = 2$
- $2^0 = 1$
- $2^1 = 2$
- $2^2 = 2 \times 2$
- $2^3 = 2^2 \times 2 = 8$
- $(2 \times 3)^2 = 2^2 \times 3^2 = 4 \times 9 = 36$ (check: $2 \times 3 = 6, 6^2 = 36$)

Rational exponents (roots)

- nth root of b: the number x such that $x^n = b$
- Written as $b^{1/n}$ or $\sqrt[n]{b}$
- Some identities (for b positive):
  - $b = (b^n)^{1/n}$
  - $b^{m/n} = (b^m)^{1/n} = \sqrt[n]{b^m}$
- Example: $\sqrt{1/(36x^2)} = \dfrac{\sqrt{1}}{\sqrt{36x^2}} = \dfrac{1}{\sqrt{36}\sqrt{x^2}} = \dfrac{1}{6x}$

Exponential function

- An important constant: e, approximately 2.718
- Useful as a base for powers
- Define $\exp(x) = e^x$
- Useful identity: $\exp(x + y) = e^{x+y} = \exp(x) \times \exp(y)$

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

- Exponents correspond to multiplication
- Addition is easier than multiplication
- Logarithms (logs) transform multiplication into addition!

Logs: definition

- Defined as the inverse operation of exponentiation
- Takes a base b and a number x
- The log of x to base b is the number y such that $b^y = x$
- Written as $\log_b(x) = y$
- Natural log: $\log_e(x)$, commonly written $\log(x)$

Logs: definition

- Undefined for $x \leq 0$
- Log is an increasing function: as x increases, $\log_b(x)$ increases
- $\log_b(b) = 1$

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: $\log_b(b^x) = x \log_b(b) = x$

Example: log identities

- Multiplication: $\log(2 \times 3) = \log(2) + \log(3) = \log(6)$
- Logs of numbers < 1 are negative: $\log(2/3) = \log(2) - \log(3) < 0$
- Power: $\log(x^2) = 2\log(x)$

Common bases, changing base

- The three most common bases: e, 10, and 2
- e — common in mathematics
- 10 — common for calculating numbers in the decimal system
- 2 — common in computer science
- Changing between bases: $\log_b(x) = \frac{\log_k(x)}{\log_k(b)}$

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

Uses of logarithms in statistics

- Transformation of the data — look at a multiplicative relationship rather than an additive relationship
- Logistic regression, Poisson regression
- For more, see BIOST 512/513!

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)