

# Notes on Logarithms

## What is a logarithm?

Statistics is often thought of as a mathematical subject. For readers of research and for users of statistical methods, however, the mathematics seldom put in much of an appearance. They are hidden down in the computer program engine room. All we actually see on deck are basic mathematical operations (we add and subtract, divide and multiply) and the occasional square root. There is only one other mathematical operation with which the user of statistics must become familiar: the logarithm. We come across logarithms in graphical presentation (logarithmic scales), in relative risks and odds ratios (standard errors and confidence intervals, logistic regression), in the transformation of data to have a Normal distribution or uniform variance, in the analysis of survival data (hazard ratios), and many more. In these notes I explain what a logarithm is, how to use logarithms, and try to demystify this most useful of mathematical tools.

I shall start with logarithms (usually shortened to 'log') to base 10. In mathematics, we write  $10^2$  to mean  $10 \times 10$ . We call this '10 to the power 2' or '10 squared'. We have  $10^2 = 10 \times 10 = 100$ . We call 2 the **logarithm of 100 to base 10** and write it as  $\log_{10}(100) = 2$ . In the same way,  $10^3 = 10 \times 10 \times 10$  is '10 to the power 3' or '10 cubed',  $10^3 = 1000$  and  $\log_{10}(1000) = 3$ .  $10^5 = 10 \times 10 \times 10 \times 10 \times 10$  is '10 to the power 5',  $10^5 = 100,000$  and  $\log_{10}(100,000) = 5$ . 10 raised to the power of the log of a number is equal to that number.  $10^1 = 10$ , so  $\log_{10}(10) = 1$ .

Before the days of electronic calculators, logarithms were used to multiply and divide large or awkward numbers. This is because when we add on the log scale we multiply on the natural scale. For example

$$\begin{aligned}\log_{10}(1000) + \log_{10}(100) &= 3 + 2 = 5 = \log_{10}(100,000) \\ 10^3 \times 10^2 &= 10^{3+2} = 10^5 \\ 1000 \times 100 &= 100,000\end{aligned}$$

Adding the log of 1000, which is 3, and the log of 100, which is 2, gives us 5, which is the log of 100,000, the multiple of 1000 and 100. So adding on the log scale is equivalent to multiplying on the natural scale.

When we subtract on the log scale we divide on the natural scale. For example

$$\begin{aligned}\log_{10}(1000) - \log_{10}(100) &= 3 - 2 = 1 = \log_{10}(10) \\ 10^3 \div 10^2 &= 10^{3-2} = 10^1 \\ 1000 \div 100 &= 10\end{aligned}$$

Subtracting the log of 100, 2, from the log of 1000, 3, gives us 1, which is the log of 10, 1000 divided by 100. So subtracting on the log scale is equivalent to dividing on the natural scale.

So far we have raised 10 to powers which are positive whole numbers, so it is very easy to see what 10 to that power means; it is that number of 10s multiplied together. It is not so easy to see what raising 10 to other powers, such as negative numbers, fractions, or zero would mean. What mathematicians do is to ask what powers other than positive whole numbers would mean if they were consistent with the definition we started with.

What is ten to the power zero? The answer is  $10^0 = 1$ , so  $\log_{10}(1) = 0$ . Why is this? Let us see what happens when we divide a number, 10 for example, by itself:

$$10 \div 10 = 1$$

$$\log_{10}(10) - \log_{10}(10) = 1 - 1 = 0 = \log_{10}(1)$$

When we subtract the log of 10, which is 1, from the log of 10, 1, the difference is zero. This must be the log of 10 divided by 10, so zero must be the logarithm of one.

So far, we have added and subtracted logarithms. If we multiply a logarithm by a number, on the natural scale we raise to the power of that number. For example:

$$3 \times \log_{10}(100) = 3 \times 2 = 6 = \log_{10}(1,000,000)$$

$$100^3 = 1,000,000.$$

If we divide a logarithm by a number, on the natural scale we take that number root. For example,  $\log_{10}(1,000)/3 = 3/3 = 1 = \log_{10}(10)$  and the cube root of 1,000 is 10, i.e.  $10 \times 10 \times 10 = 1,000$ . Note that we are multiplying and dividing a logarithm by a plain number, not by another logarithm.

### Logarithms which are not whole numbers

Logarithms do not have to be whole numbers. For example, 0.5 (or  $\frac{1}{2}$ ) is the logarithm of the square root of 10. We have  $10^{0.5} = 10^{\frac{1}{2}} = \sqrt{10} = 3.16228$ . We know this because

$$10^{\frac{1}{2}} \times 10^{\frac{1}{2}} = 10^{\frac{1}{2} + \frac{1}{2}} = 10^1 = 10.$$

We do not know what 10 to the power  $\frac{1}{2}$  means. We do know that if we multiply 10 to the power  $\frac{1}{2}$  by 10 to the power  $\frac{1}{2}$ , we will have 10 to the power  $\frac{1}{2} + \frac{1}{2} = 1$ . So 10 to the power  $\frac{1}{2}$  multiplied by itself is equal to 10 and  $10^{\frac{1}{2}}$  must be the square root of 10. Hence  $\frac{1}{2}$  is the log to base 10 of the square root of 10.

Logarithms which are not whole numbers are the logs of numbers which cannot be written as 1 and a string of zeros. For example the log of 2 is 0.30103 and the log of 5 is 0.69897. Of course, these add to 1, the log of 10, because  $2 \times 5 = 10$ :

$$0.30103 + 0.69897 = 1.0000$$

Negative logarithms are the logs of numbers less than one. For example, the log of 0.1 is  $-1$ . This must be the case, because 0.1 is one divided by 10:

$$1 \div 10 = 0.1$$

$$\log_{10}(1) - \log_{10}(10) = 0 - 1 = -1 = \log_{10}(0.1)$$

In the same way, the log of  $\frac{1}{2}$  is minus the log of 2:  $\log_{10}(\frac{1}{2}) = -0.30103$ . Again, this is consistent with everything else. For example, if we multiply 2 by  $\frac{1}{2}$  we will get one:

$$\log_{10}(2) + \log_{10}(\frac{1}{2}) = 0.30103 - 0.30103 = 0 = \log_{10}(1)$$

$$2 \times \frac{1}{2} = 1$$

What is  $\log_{10}(0)$ ? It does not exist. There is no power to which we can raise 10 to give zero. To get a multiple equal to zero, one of the numbers multiplied must equal zero. As we take the logs of smaller and smaller numbers, the logs are larger and larger negative numbers. For example,  $\log_{10}(0.0000000001) = -10$ . We say that the log of a number tends towards minus infinity (written as  $-\infty$ ) as the number tends

towards zero. The logarithms of negative numbers do not exist, either. We can only use logarithms for positive numbers.

### The logarithmic curve and logarithmic scale

Figure 1 shows the curve representing the logarithm to base 10. The curve starts off at the bottom of the vertical scale just right of zero on the horizontal scale, coming up from minus infinity as the log of zero, if we were able to get it on the paper. It goes through the point defined by 1 on the horizontal axis and 0 on the vertical axis, and continues to rise but less and less steeply, going through the points defined by 10 and 1 and by 100 and 2.

The right hand vertical axis of Figure 1 shows the variable on a logarithmic scale. The scale is marked in unequal divisions which correspond to the logarithms of the numbers printed. On this scale, the distance between 1 and 10 is the same as the distance between 10 and 100. So equal distances mean equal ratios ( $10/1 = 100/10$ ) rather than equal differences, as on a linear or natural scale.

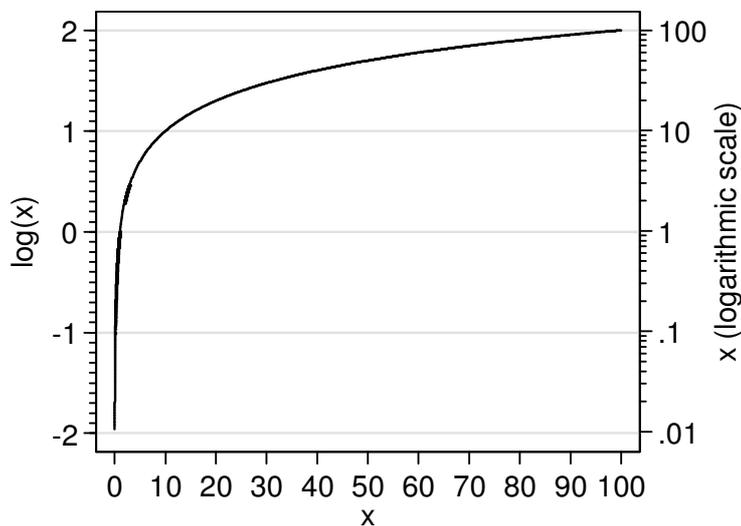


Figure 1 Graph of the logarithm to base 10, with a logarithmic scale on the right side.

Showing data on a logarithmic scale can often show us details which are obscured on the natural scale. For example, Figure 2 shows Prostate Specific Antigen (PSA) for three groups of subjects, men with prostate cancer, with inflammation of the prostate (prostatitis) and normal controls. We can see that a few of the cancer patients have high PSA, but we can see very little else. Figure 3 shows the log of PSA. A lot more detail is clear, including the huge overlap between the three groups.

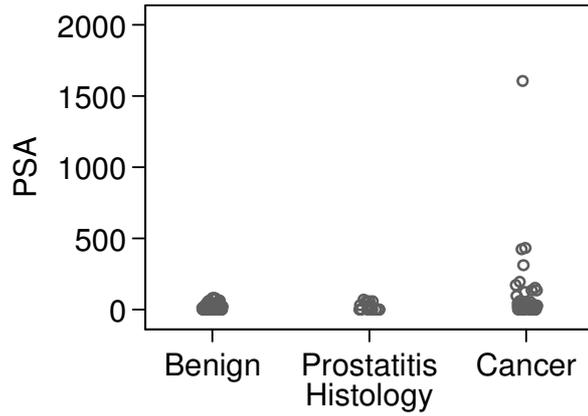


Figure 2. PSA for three groups of subjects, natural scale.

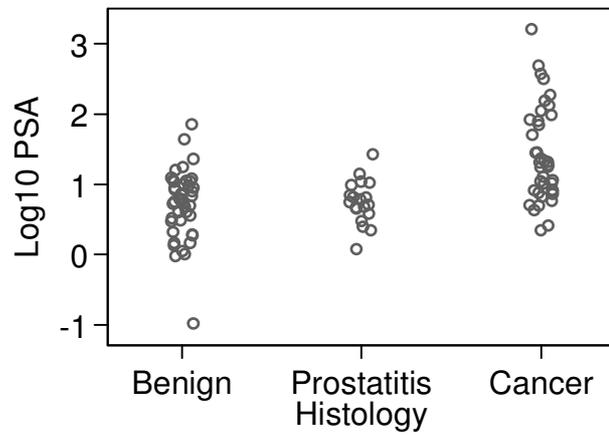


Figure 3.  $\text{Log}_{10}(\text{PSA})$  for three groups of subjects.

The units in which  $\text{log}_{10}(\text{PSA})$  is measured may not be easily understood by those who use PSA measurements. Instead, we can put the original units from Figure 2 onto the graph shown in Figure 3 by means of a logarithmic scale. Figure 4 shows the PSA on a logarithmic scale, the structure revealed by the logarithm is shown but in the original units.

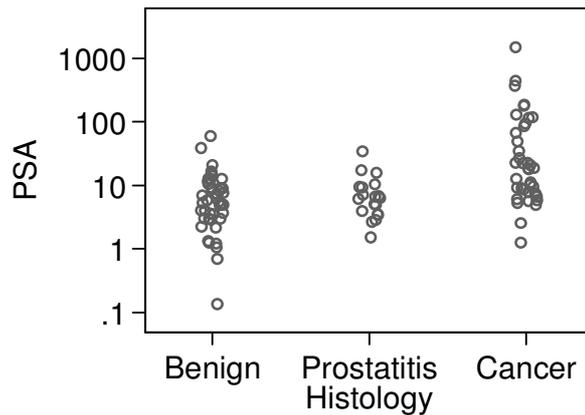


Figure 4. PSA for three groups of subjects, logarithmic scale.

### Natural logarithms and base ‘e’

An early use of logarithms was to multiply or divide large numbers, to raise numbers to powers, etc. For these calculations, 10 was the obvious base to use, because our number system uses base 10, i.e. we count in tens. We are so used to this that we might think it is somehow inevitable, but it happened because we have ten fingers and thumbs on our hands. If we had had twelve digits instead, we would have counted to base 12, which would have made a lot of arithmetic much easier. Other bases have been used. The ancient Babylonians, for example, are said to have counted to base 60, though perhaps only a few of them did much counting at all.

Base 10 for logarithms was chosen for convenience in arithmetic, but it was a choice, it was not the only possible base. Logarithms to the base 10 are also called **common logarithms**, the logarithms for everyone to use.

Mathematicians also find it convenient to use a different base, called ‘e’, to give **natural logarithms**. The symbol ‘e’ represents a number which cannot be written down exactly, like  $\pi$  (pi, the ratio of the circumference of a circle to its diameter). In decimals,  $e = 2.718281 \dots$  and this goes on and on indefinitely, just like  $\pi$ . We use this base because the slope of the curve  $y = \log_{10}(x)$  is  $\log_{10}(e)/x$ . The slope of the curve  $y = \log_e(x)$  is  $1/x$ . It displays a rather breathtaking insouciance to call ‘natural’ the use of a number which you cannot even write down and which has to be labelled by a letter, but that is what we do. Using natural logs avoids awkward constants in formulae and as long as we are not trying to use them to do calculations, it makes life much easier. When you see ‘log’ written in statistics, it is the natural log unless we specify something else.

Logs to base e are sometimes written as ‘ln’ rather than ‘ $\log_e$ ’. On calculators, the button for natural logs is usually labelled ‘ln’ or ‘ln(x)’. The button labelled ‘log’ or ‘log(x)’ usually does logs to the base 10. If in doubt, try putting in 10 and pressing the log button. As we have seen,  $\log_{10}(10) = 1$ , whereas  $\log_e(10) = 2.3026$ .

### Antilogarithms

The antilogarithm is the opposite of the logarithm. If we start with a logarithm, the **antilogarithm** or **antilog** is the number of which this is the logarithm. Hence the

antilog to base 10 of 2 is 100, because 2 is the log to base 10 of 100. To convert from logarithms to the natural scale, we antilog:

$$\text{antilog}_{10}(2) = 10^2 = 100$$

We usually write this as  $10^2$  rather than  $\text{antilog}_{10}(2)$ . On a calculator, the antilog key for base 10 is usually labelled ' $10^x$ '.

To antilog from logs to base e on a calculator, use the key labelled ' $e^x$ ' or ' $\text{exp}(x)$ '. Here ' $\text{exp}$ ' is short for 'exponential'. This is another word for 'power' in the sense of 'raised to the power of' and the mathematical function which is the opposite of the log to base e, the antilog, is called the exponential function. So 'e' is for 'exponential'.

We also use the term 'exponential' to describe a way of writing down very large and very small numbers. Suppose we want to write the number 1,234,000,000,000. Now this is equal to  $1.234 \times 1,000,000,000,000$ . We can write this as  $1.234 \times 10^{12}$ . Similarly, we can write 0.000,000,000,001234 as  $1.234 \times 0.000,000,000,001 = 1.234 \times 10^{-12}$ . Computers print these numbers out as 1.234E12 and 1.234E-12. This makes things nice and compact on the screen or printout, but can be very confusing to the occasional user of numerical software.

Martin Bland  
1 March 2006