**Ghement Statistical Consulting Company Ltd. © 2013**

**Dealing with very small numbers (presented in scientific notation)**

**when constructing R plots**

Very small numbers (e.g., 0.0000015) often appear on the vertical axis of an R plot and R doesn’t do a good job at representing such numbers. To illustrate this, let us generate some data and use these data to construct a density histogram.

set.seed(1)

x <- seq(1,10,length=100)

y <- x\*100000 + rnorm(n=length(x),mean=0, sd=50000)

hist(y,prob=TRUE)



If we focus our attention exclusively on the numbers displayed on the vertical axis of the histogram, we notice two things:

1. The numbers are displayed using scientific notation;
2. The numbers are very small.

For instance, the number shown at the top of the vertical axis of the histogram is 1.5e-06 which is scientific notation for the very small number 0.0000015. Now it is clear that 1.5e-06 is R’s way of representing 0.0000015 and that 1.5e-06 is scientific notation for $1.5∙10^{-6}$.

How can we force R to abandon the scientific notation used for the vertical axis (e.g.,1.5e-06) and replace it instead with the regular notation (e.g., 0.0000015)?

The R code below provides the answer to this question. The code uses some clever R tricks such as:

1. Plotting the density histogram without an y-axis and y-axis label using the hist() function;
2. Extracting the underlying densities from a subsequent call to hist() which uses the option plot=FALSE;
3. Plotting the y-axis using “pretty” values for the underlying densities, which are formatted so they are not represented in a scientific notation;
4. Adding a label for the y-axis with the mtext() function.

par(mar=c(4,7,2,2))

hist(y,prob=TRUE, yaxt="n", ylab="")

h <- hist(y,plot=FALSE)

dens <- h$density

axis(2,at=pretty(dens), labels=format(pretty(dens), scientific=FALSE), las=1)

mtext(text="y-axis label", side=2, line=6)



If we now examine the horizontal axis of the histogram, we realize that we should also change the formatting of the values represented on this axis. These values are currently displayed in scientific notation and are very large. We will therefore force R to replace these values (e.g., 1e+06) with nicer ones (e.g., 1,000,000). The R code we need to implement this change and its corresponding output are given below:

par(mar=c(4,7,2,2))

hist(y,prob=TRUE, xaxt="n", yaxt="n", xlab="", ylab="")

h <- hist(y,plot=FALSE)

dens <- h$density

axis(1,at=pretty(y), labels=format(pretty(y), big.mark="," ,scientific=FALSE),

 las=1)

axis(2,at=pretty(dens), labels=format(pretty(dens), scientific=FALSE), las=1)

mtext(text="x-axis label", side=1, line=2.5)

mtext(text="y-axis label", side=2, line=6)



The density histogram looks more reasonable as the values displayed on both axes can be easily read.

**Comment 1:**

What is the meaning of the various options used in the R command below?

axis(2,at=pretty(dens), labels=format(pretty(dens), scientific=FALSE), las=1)

When we use axis(2, ...), we are instructing R to draw a vertical axis on a plot which lacks a vertical axis.

When we use axis(2,at=pretty(dens),...), we are telling R to place tick marks on the vertical axis at locations given by the values stored in the vector pretty(dens).

When we use axis(2,at=pretty(dens), labels=format(pretty(dens), scientific=FALSE),...), we are asking R to label the values placed at specified tick marks using a non-scientific notation.

In other words, the option at= controls where the values on the y-axis are placed, whereas the option labels= controls how these values are formatted and displayed.

When we use axis(2,at=pretty(dens), labels=format(pretty(dens), scientific=FALSE), las=1), R knows that it should position the values shown on the y-axis so that they are perpendicular to this axis.

**Comment 2:**

The function pretty() is used to produce a set of nicer values for display on the axes of a particular R plot. To understand how this function works, let’s try to prettify the numbers 2.07, 3.78 and 4.95:

x <- c(2.07, 3.78, 4.95)

pretty(x)

The outcome of this pretty-fication is:

pretty(x)

[1] 2.0 2.5 3.0 3.5 4.0 4.5 5.0

According to the R help file for the pretty() function, this function computes a sequence of about n+1 equally spaced ‘round’ values which cover the range of the values in x. The values are chosen so that they are 1, 2 or 5 times a power of 10.

**Comment 3:**

The function format() can be used to control the format of a number (e.g., scientific versus non-scientific).

As an example, if we take the number 1e+03 (which is scientific notation for 1000), we can strip it of its scientific notation using the R command below:

format(1e+3, scientific=FALSE)

The result of this operation is a character value "1000". We can convert this character value into a numeric value with the R command as.numeric(format(1e+3, scientific=FALSE)).

Conversely, we can use the R command format(1000, scientific=TRUE) to convert the number 1000 into its scientific version 1e+03.