Chapter 11 - The Lognormal Distribution and the Geometirc Mean

Laura Saba

July 12, 2017

What this chapter covers

- The origin of a lognormal distribution
- How to analyze lognormal data
- Geometric mean
- Common mistakes: Lognormal distribution

THE ORIGIN OF A LOGNORMAL DISTRIBUTION

- When many independent factors influence a measured value in an additive fashion, the measured values are likely to follow a *normal distribution*.
- When many independent factos influence a measured value in a multiplicative fashion, the measured values are likely to follow a *lognormal distribution*.

THE ORIGIN OF A LOGNORMAL DISTRIBUTION

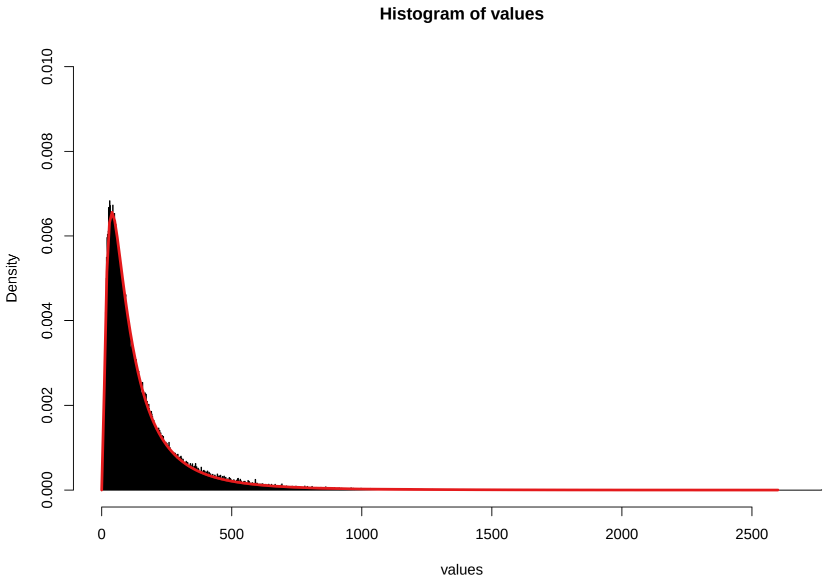
```r
true_value = 100
n = 100000
e1 = runif(n,1,2)^(sample(c(-1,1),n,replace=TRUE))
e2 = runif(n,1,2)^(sample(c(-1,1),n,replace=TRUE))
e3 = runif(n,1,2)^(sample(c(-1,1),n,replace=TRUE))
e4 = runif(n,1,2)^(sample(c(-1,1),n,replace=TRUE))
e5 = runif(n,1,2)^(sample(c(-1,1),n,replace=TRUE))
values = true_value*e1*e2*e3*e4*e5
head(round(e1,2))
```

```
## [1] 1.05 1.28 0.64 1.80 1.46 1.12
```

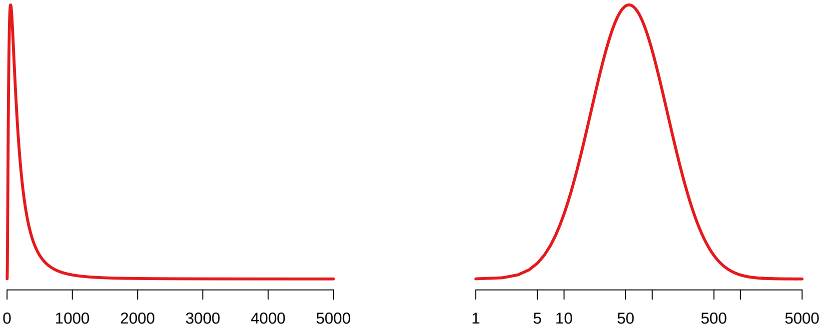
```r
head(round(values,2))
```

```
## [1] 140.58 118.31  13.72 184.14  20.92 136.57
```

Result of many random factors contributing

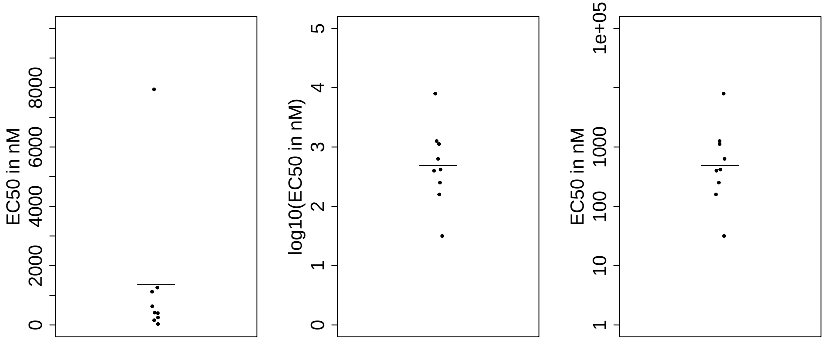


Plotting on the logarithmic scale



Example Data - Figure 11.1

```
EC50 = 10^c(1.5,2.2,2.4,2.6,2.62,2.8,3.1,3.05,3.9)
```



Why logarithms?

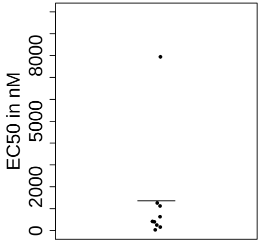
Properties of logarithms:

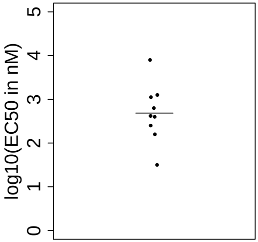
- $\log(ab) = \log(a) + \log(b)$
- Logarithms convert multiplicative scatter (lognormal distribution) to additive scatter (Gaussian)

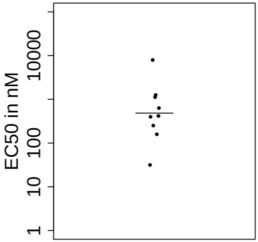
HOW TO ANALYZE LOGNORMAL DATA

- Typically, the logged data is used for statistical analyses
- Results (figures, tables, etc.) are transformed to the antilog to keep the magnitude of estimates interpretable

GEOMETRIC MEAN







- For the 1st, the mean is indicated by the line. This appears to be a bad summary of the central tendency of the data because only one value is greater than the mean and the remaining values are below the mean.
- For the 2nd and 3rd graphs, the **geometric mean** is indicated by the line. This is a much better summary of the central tendency of the data because close to equal numbers of values are above and below the line.

COMMON MISTAKES: LOGNORMAL DISTRIBUTION

1. Being inconsistent with the use of common and natural logs
    - the same base should be used for initially converting values to their logarithm and when taking the antilogarithm of the mean of the logged values
2. Converting the data to logarithms when some values are zero or negative
    - taking the log of zero will give a negative infinity value in R
    - taking the log of a negative number will give a warning and NaN values in R

What did we learn

- Lognormal distributions are very common in many fields of science.
- Lognormal distributions arise when multiple random factors are multiplied together to determine the value. In contrast, Gaussian distributions arise when multiple random factors are added together.
- Lognormal distributions have a long right tail (ie skewed right).
- In most cases, the best way to analyze lognormal data is to take the logarithm of each value and then analyze those logarithms.