Plant Protection and Data Science - The Normal Distribution is the Log-Normal Distribution

Eckhard Limpert* and Werner A Stahel
Seminar for Statistics, ETH Zurich, CH-8092 Zurich, Switzerland

1. Chances

Most natural laws are based on multiplication, division, and powers. This leads to the experience that the distribution which most often describes observed or measured data well is the log-normal or multiplicative normal distribution, which therefore deserves the name of "the normal distribution", rather than the established, "additive" Gaussian or normal distribution. Using this fact as a starting point in data analysis bears great chances. This is true for a variety of applications across plant protection and most sciences[1-3].

2. Nature multiplies

Multiplication of cells 1-2-4-8-16-…

More general, growth, of plants, capital etc: $s_t = s_0 \cdot k^t$

Life is chemistry, e.g., reaction velocity, $v = k \cdot [A] \cdot [B]$

mass action, $A+B \rightleftarrows C+D$: $K_c = ([A] \cdot [B]) / ([C] \cdot [D])$

and physics, *e.g.,* osmosis, gas laws $p \cdot v = n \cdot R \cdot T$

Finally, Einstein's form assumed to be the most well known is based on multiplication, too: $E = m \cdot c^2$

3. Log- or multiplicative normal distributions

Moreover, the multiplicative version of the well known Central Limit Theorem gives a reason for the basic role of the multiplicative normal distribution: If a number of relatively small variations of single variables in the formulas are combined by multiplications or divisions, this leads to an approximate multiplicative normal distribution.

In addition to the natural laws indicated, a number of phenomena such as health risk (Hattis et al., 1999 cf[2]) permeability and solute mobility (e.g. Baur, 1997, cf[1,2]) or many further mechanisms (Koch 1966, cf[1]) leed to log-normal distributions of data.

In line with this, examples fitting log-normal distributions are ubiquitously found across the sciences. In plant protection, they relate to various aspects and interactions within the disease triangle composed of plants, pests and pathogens, and the environment.

```
                THE
             NORMAL LAW
           OF ORIGINAL DATA
          IS THE LOG-NORMAL
         DISTRIBUTION ◊ IT STANDS
        OUT IN THE EXPERIENCE OF
       MANKIND AS ONE OF THE BROADEST
      GENERALIZATIONS OF NATURAL PHILOSOPHY ◊
     IT SERVES AS THE GUIDING INSTRUMENT IN
    RESEARCHES IN THE PHYSICAL AND SOCIAL SCIENCES
   AND IN MEDICINE, AGRICULTURE, AND ENGINEERING ◊
  IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT
```

4. The normal distribution is the log-normal distribution. Starting from the impressive anthem and monument W.J. Youden designed in honor of the normal distribution, we changed some words in the beginning only to adapt it to the log-normal. - The shape corresponds to a shape parameter s* of 1.5. Then, 68% of the data are in the range from the multiplicative mean xbar* times/divide ($^{x}/$) 1.5, with 95% in the range $^{x}/ 1.5^2$. Interestingly, this s* is quite typical for latent periods of infectious diseases. – Now, how would be ranges for further interactions and s*'s?

5. According to this insight, analyses should be based on the multiplicative normal distribution in order to improve the quality and efficiency of results. If humans or animals are involved, this even is an ethical imperative. For further results see our poster on "Reconsidering the normal distribution – benefits from replacing ± by $^{x}/$ ".

*Present Address: ELI-o-Research, Scheuchzerstr. 210, CH-8057 Zürich; +41-76-369 2132 (handy); eckhard.limpert@bluewin.ch

References

1. Limpert E, Stahel WA, 2011, Problems with Using the Normal Distribution – and Ways to Improve Quality and Efficiency of Data Analysis. PLoS ONE 6(7):e21403. doi:10.1371/journal.pone.0021403
2. Limpert E, Stahel WA, Abbt M., 2001, Log-normal distributions across the sciences - keys and clues. BioScience 51, 341-352.
3. Stahel WA, Limpert E, The normal distribution is the log-normal distribution. Talk Leibniz-Inst. Magdeburg, Dec 2, 2014, http://stat.ethz.ch/~stahel/talks/lognormal.pdf