Properties of log transformations and backtransformations

Example uses hamburger bacterial concentration (data in hamburger.csv)
  Reminder on study design:
    2 treatments (active, control), 6 observations per treatment
    One unusual cfu that isn't so unusual on the log scale
  Summary statistics:

| | Active | Control | Diff |
|---|---|---|---|
| average cfu | 0.11 | 0.72 | |
| median cfu | 0.11 | 0.44 | |
| average log cfu | -2.535 | -0.769 | 1.76 |

Three properties of logs, means and medians

1. exp( average $\log Y$) $\neq$ average$Y$
   Control group: exp(-0.769) = 0.46
     not close to the data mean, 0.72
   Details on why this is the case:
     model: $\log Y \sim N(\mu_l, \sigma_l^2)$
       Math Stat fact: mean $Y = \exp(\mu_l + \sigma_l^2/2)$
     $\overline{\log Y}$ estimates $\mu_l$
     $\exp \overline{\log Y}$ estimates $\exp \mu_l$
       but mean $Y = \exp(\mu_l + \sigma_l^2/2) =$ (what $\exp \overline{\log Y}$ estimates) $\times$ ($\exp \sigma_l^2/2$)
       $\exp \overline{\log Y}$ systematically smaller than the mean
       How much smaller depends on $\sigma_l$

2. Back-transforming average $\log Y$ estimates the median of $Y$, not the mean
   Control group: $\exp -0.769 = 0.46$, close to the data-based median = 0.44
   The details:
     median ( $\log Y$) = log(median $Y$)
     When $\log Y$ approximately symmetrical (i.e., not skewed), median = mean
     When exactly symmetrical:
       median ( $\log Y$) = mean ($\log Y$), estimated by average ($\log Y$)
     exp average $\log Y = \exp \log($ median $Y) =$ median $Y$   Two different ways (data median, back transformation) to estimate the same parameter
     Two different estimators usually give different estimates
       Lots of theory on which is better and under what conditions

3. Back transforming the difference of average $\log Y$ estimates the ratio of medians
   diff = average $\log Y_1$ - average $\log Y_2$ = median $\log Y_1$ - median $\log Y_2$
   = log(median $Y_1$) $-$ log(median $Y_2$)
   = log (median $Y_1$/median $Y_2$)
   so exp diff = exp log (median $Y_1$/median $Y_2$)
   = median $Y_1$/median $Y_2$