Hodges–Lehmann median differences between exponential subpopulations

Roger B. Newson

12 October, 2008

1 Formulas

Suppose that $Y_{\rm min}$ and $Y_{\rm maj}$ are scalar random variables, sampled independently from 2 exponential subpopulations, with means (and therefore also standard deviations) equal to $\sigma_{\rm min}$ and $\sigma_{\rm maj}$ respectively, where $\sigma_{\rm min} \le \sigma_{\rm maj}$. Given $q \in (0,1)$, we aim to define formulas for the $100q$th percentile differences $\xi_q(Y_{\rm maj} - Y_{\rm min})$, defined as solutions in θ to the equation

$$\Pr\{Y_{\rm maj} - Y_{\rm min} \le \theta\} = q. \tag{1}$$

In particular, we aim to define a formula for $\xi_{0.5}(Y_{\rm maj} - Y_{\rm min})$, known as the Hodges–Lehmann median difference between $Y_{\rm maj}$ and $Y_{\rm min}$. (See Hodges and Lehmann (1963) and Lehmann (1963).)

The variables $Y_{\rm maj}$ and $Y_{\rm min}$ have constant hazard rates $\sigma_{\rm maj}^{-1}$ and $\sigma_{\rm min}^{-1}$, respectively. It follows that

$$\begin{array}{rcl} \Pr\{Y_{\rm maj} - Y_{\rm min} < 0\} & = & \sigma_{\rm maj}^{-1}/(\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}), \\ \Pr\{Y_{\rm maj} - Y_{\rm min} > 0\} & = & \sigma_{\rm min}^{-1}/(\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}), \end{array} \tag{2}$$

and that the conditional distribution of $Y_{\rm min} - Y_{\rm maj}$ given that $Y_{\rm maj} < Y_{\rm min}$, and the conditional distribution of $Y_{\rm maj} - Y_{\rm min}$ given that $Y_{\rm min} < Y_{\rm maj}$, are both exponential, with means $\sigma_{\rm min}$ and $\sigma_{\rm maj}$, respectively. It follows that, for any real θ,

$$\Pr\{Y_{\rm maj} - Y_{\rm min} \le \theta\} = \begin{cases} \frac{\sigma_{\rm maj}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} \exp(\theta/\sigma_{\rm min}), & \text{if } \theta < 0, \\ \frac{\sigma_{\rm maj}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} + \frac{\sigma_{\rm min}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} \left[1 - \exp(-\theta/\sigma_{\rm maj})\right], & \text{if } \theta \ge 0. \end{cases} \tag{3}$$

Therefore, given $q \in (0,1)$, the $100q$th percentile difference $\xi_q(Y_{\rm maj} - Y_{\rm min})$ is a solution in θ to the equation

$$q = \begin{cases} \frac{\sigma_{\rm maj}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} \exp(\theta/\sigma_{\rm min}), & \text{if } q < \sigma_{\rm maj}^{-1}/(\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}), \\ \frac{\sigma_{\rm maj}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} + \frac{\sigma_{\rm min}^{-1}}{\sigma_{\rm maj}^{-1}+\sigma_{\rm min}^{-1}} \left[1 - \exp(-\theta/\sigma_{\rm maj})\right], & \text{if } q \ge \sigma_{\rm maj}^{-1}/(\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}). \end{cases} \tag{4}$$

(This is a consequence of (1), (2), (3), and the fact that a cumulative distribution function is monotonically nondecreasing.) The Hodges–Lehmann median difference $\xi_{0.5}(Y_{\rm maj} - Y_{\rm min})$ corresponds to the case where $q = 0.5 = 1 - q$, and is a solution to the second case of (4) (because $\sigma_{\rm maj}^{-1} \le \sigma_{\rm min}^{-1}$). In this case, the equation to solve in θ is

$$\frac{\sigma_{\rm maj}^{-1}}{\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}} + \frac{\sigma_{\rm min}^{-1}}{\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}} \left[1 - \exp(-\theta/\sigma_{\rm maj})\right] = \frac{\sigma_{\rm min}^{-1}}{\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}} \exp(-\theta/\sigma_{\rm maj}), \tag{5}$$

or, equivalently,

$$\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1} = 2\sigma_{\rm min}^{-1} \exp(-\theta/\sigma_{\rm maj}), \tag{6}$$

implying that the solution is

$$\xi_{0.5}(Y_{\rm maj} - Y_{\rm min}) = -\sigma_{\rm maj} \ln\left[\frac{\sigma_{\rm maj}^{-1} + \sigma_{\rm min}^{-1}}{2\sigma_{\rm min}^{-1}}\right] = -\sigma_{\rm maj} \ln\left[\frac{\sigma_{\rm min}/\sigma_{\rm maj} + 1}{2}\right]. \tag{7}$$

Similarly, the Hodges–Lehmann median difference between $Y_{\rm min}$ and $Y_{\rm maj}$ can be defined as

$$\xi_{0.5}(Y_{\rm min} - Y_{\rm maj}) = \sigma_{\rm maj} \ln\left[\frac{\sigma_{\rm min}/\sigma_{\rm maj} + 1}{2}\right]. \tag{8}$$

1.1 Median differences and differences between medians

Note that, in general, the Hodges–Lehmann median difference is *not* equal to the difference between the two subpopulation medians, or to the difference between the two subpopulation means. The median (or half–life) of an exponential distribution is equal to its mean multiplied by ln(2). Therefore, the difference between the medians of $Y_{\rm maj}$ and $Y_{\rm min}$ is defined as

$$\xi_{0.5}(Y_{\rm maj}) - \xi_{0.5}(Y_{\rm min}) = (\sigma_{\rm maj} - \sigma_{\rm min})\ln(2)\,. \tag{9}$$

And the mean difference (which *is* equal to the difference between the means) is defined as

$$E(Y_{\rm maj} - Y_{\rm min}) = E(Y_{\rm maj}) - E(Y_{\rm min}) = \sigma_{\rm maj} - \sigma_{\rm min}\,. \tag{10}$$

All of these differences are equal to zero when $\sigma_{\rm maj} = \sigma_{\rm min}$. However, they are unequal and nonzero when $\sigma_{\rm maj} > \sigma_{\rm min}$.

For example, suppose (without loss of generality) that $\sigma_{\rm min} = 1$ and $\sigma_{\rm maj} = \sigma_{\rm rat} \geq 1$. Then the median difference, the difference between medians, and the mean difference are given by

$$\begin{aligned} \xi_{0.5}(Y_{\rm maj} - Y_{\rm min}) &= -\sigma_{\rm rat}\ln\left[(\sigma_{\rm rat}^{-1}+1)/2\right], \\ \xi_{0.5}(Y_{\rm maj}) - \xi_{0.5}(Y_{\rm min}) &= (\sigma_{\rm rat}-1)\ln(2), \\ E(Y_{\rm maj} - Y_{\rm min}) &= \sigma_{\rm rat} - 1. \end{aligned} \tag{11}$$

These are identically zero if $\sigma_{\rm rat} = 1$. Their derivatives with respect to $\sigma_{\rm rat}$ are given by

$$\begin{aligned} \frac{\partial}{\partial\sigma_{\rm rat}}\xi_{0.5}(Y_{\rm maj} - Y_{\rm min}) &= \ln(2) - \ln(\sigma_{\rm rat}^{-1}+1) + (\sigma_{\rm rat}+1)^{-1}, \\ \frac{\partial}{\partial\sigma_{\rm rat}}\left[\xi_{0.5}(Y_{\rm maj}) - \xi_{0.5}(Y_{\rm min})\right] &= \ln(2), \\ \frac{\partial}{\partial\sigma_{\rm rat}}E(Y_{\rm maj} - Y_{\rm min}) &= 1. \end{aligned} \tag{12}$$

If $\sigma_{\rm rat} = 1$, then these derivatives are 0.5, ln(2) and 1, respectively. Therefore, for an open interval of $\sigma_{\rm rat}$ values immediately to the right of 1, we have the inequality

$$\xi_{0.5}(Y_{\rm maj} - Y_{\rm min}) < \xi_{0.5}(Y_{\rm maj}) - \xi_{0.5}(Y_{\rm min}) < E(Y_{\rm maj} - Y_{\rm min})\,. \tag{13}$$

The second derivative of the Hodges–Lehmann median difference with respect to $\sigma_{\rm rat}$ is

$$\frac{\partial^2}{\partial\sigma_{\rm rat}^2}\xi_{0.5}(Y_{\rm maj} - Y_{\rm min}) = \frac{1}{\sigma_{\rm rat}(1+\sigma_{\rm rat})} - \frac{1}{(1+\sigma_{\rm rat})^2}\,, \tag{14}$$

which is positive and monotonically decreasing in $\sigma_{\rm rat}$ for $\sigma_{\rm rat} \geq 1$, is equal to 0.25 if $\sigma_{\rm rat} = 1$, and tends to zero in the limit as $\sigma_{\rm rat} \to \infty$. Therefore, the median difference is asymptotically linear in $\sigma_{\rm rat}$, with limiting slope ln(2), and the difference between medians and the mean difference are linear in $\sigma_{\rm rat}$, with slopes ln(2) and 1, respectively. And the inequality (13) holds for all $\sigma_{\rm rat} > 1$, with the differences between the 3 functionals increasing in magnitude with $\sigma_{\rm rat}$. Figure 1 illustrates the 3 functionals plotted against $\sigma_{\rm rat}$, over the domain $1 \leq \sigma_{\rm rat} \leq 4$.

However, although the median difference and the difference between medians do not converge in difference, they do converge in ratio. For $\sigma_{\rm rat} > 1$, the ratio between the difference between the medians and the median difference is

$$\frac{\xi_{0.5}(Y_{\rm maj}) - \xi_{0.5}(Y_{\rm min})}{\xi_{0.5}(Y_{\rm maj} - Y_{\rm min})} = \frac{\ln(2)}{\ln(2) - \ln(\sigma_{\rm rat}^{-1}+1)}(1-\sigma_{\rm rat}^{-1})\,. \tag{15}$$

This is decreasing in $\sigma_{\rm rat}$ where $\sigma_{\rm rat} > 1$, because the numerator and denominator are both zero when $\sigma_{\rm rat} = 1$, and the derivative of the numerator is constant in $\sigma_{\rm rat}$, and the derivative of the denominator is increasing in $\sigma_{\rm rat}$. It tends to $2\ln(2)$ in the limit as $\sigma_{\rm rat} \to 1$ (by (12) and L'Hospital's rule), and tends to 1 in the limit as $\sigma_{\rm rat} \to \infty$. Therefore, the median difference and the difference between the medians converge in ratio (but *not* in difference), as the ratio between the means becomes very large. Figure 2 shows the ratio as a function of $\sigma_{\rm rat}$ over the domain $1 < \sigma_{\rm rat} \leq 4$.

2 References

Hodges JL, Lehmann EL. Estimates of location based on rank tests. *The Annals of Mathematical Statistics* 1963; **34(2)**: 598–611.

Lehmann EL. Nonparametric confidence intervals for a shift parameter. *The Annals of Mathematical Statistics* 1963; **34(4)**: 1507–1512.

Figure 1: Mean difference, difference between medians, and median difference as functions of σ_{rat}.

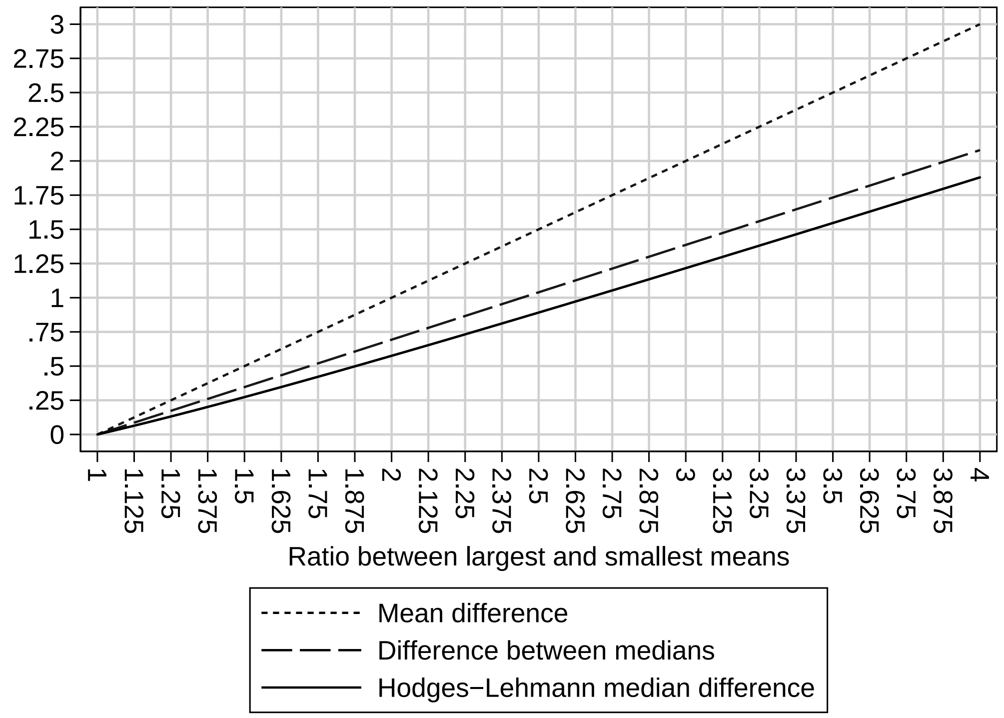

Figure 2: Ratio between difference between medians and median difference as a function of σ_{rat}.