First we consider the following problem: given a one-dimensional interval-valued dataset \([x_1, \tilde{x}_1], \ldots, [x_n, \tilde{x}_n]\) and a statistic \(S\), the task is to determine tight upper and lower bounds for the value of \(S\) over the dataset. We discuss some particular statistics such as sample variance \(\hat{\sigma}^2\) or coefficient of variation \(t = \hat{\mu}/\hat{\sigma}\). We present some examples when the computation of the bounds can be performed efficiently (in polynomial time) and when the problem is computationally hard (computable in exponential time only, unless \(P = NP\)). The complexity-theoretic results are sometimes surprising: although the statistics \(|t|\) and \(t\) are essentially the same, their behavior over interval-valued datasets differs substantially from the complexity-theoretic viewpoint. We also strengthen some results and show that not only exact computation, but also a “reasonable” approximation is computationally intractable.

Then we consider linear regression models with interval-valued data (we discuss various cases, such as interval-valued dependent variable and both interval-valued dependent and independent variables). We show under which conditions the computation of tight bounds of minimum \(L_p\)-norm estimators (such as OLS, GLS, LAD or Chebyshev) and their associated residual loss functions are computable efficiently and when we must face NP-hard problems.

Keywords: interval data; computational complexity; inapproximability.