

## How to Test Hypotheses When Exact Values are Replaced by Intervals to Protect Privacy: Case of t-Tests

V Kreinovich<sup>1</sup>, C Servin<sup>2</sup>

<sup>1</sup> *University of Texas at El Paso, El Paso, Texas*

<sup>2</sup> *El Paso Community College, El Paso, Texas*

Biomedical researchers continuously look for possible relations between relevant quantities, relations which may help in preventing and curing diseases. Once a hypothesis is made about such a relation, it is necessary to test whether it is confirmed by the data. For such hypothesis testing, t-tests are most widely used. For example, a t-test can check, based on two samples, whether it is possible that they come from distributions with the same mean – e.g., whether the average blood pressure after a proposed treatment is the same as before or it is provably smaller – meaning that the tested treatment works.

In traditional statistics, we assume that we know the exact values of the corresponding quantities. In biomedical research, however, it is important to preserve patients' privacy and confidentiality – and, knowing the exact values of all relevant parameters, one can uniquely identify the patient. One of the most efficient ways to preserve privacy is thus to replace the exact values with intervals containing such values. For example, instead of the exact age – which can uniquely identify the patient – we only store an interval containing this age: between 20 and 30, or between 30 and 40, etc.

Different values from the corresponding intervals lead, in general, to different values of the corresponding statistic. In such situations, to make sure that the data confirms the given hypothesis, we need to check that the corresponding statistic is within the desired interval for all possible values of the input quantities. In other words, we need to make sure that the whole range of possible values of the corresponding statistic is inside the desired interval. Computing this interval is, in general, NP-hard. In this talk, we provide efficient algorithms for computing t-tests under privacy-motivated interval uncertainty.