



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Seminar

CLOSED TESTING FOR SELECTIVE INFERENCE

Jelle J. Goeman

Leiden University Medical Center

**October 26, 2018 | 3.00 p.m. | Aula Cucconi
Campus S. Caterina**

Abstract : www.stat.unipd.it/fare-ricerca/seminari

CLOSED TESTING FOR SELECTIVE INFERENCE

Jelle Jurjen Goeman

Professor of Biostatistics | Department of Biomedical Data Sciences | Leiden University Medical Center, The Netherlands

In modern high-dimensional data the resolution of measurement is often different from the resolution of inference. In genomics we measure signal at genomic loci, but we are interested in inferring about signal in genomic regions. In neuroimaging we measure signal at the level of voxels, but we are interested in inference about signal in brain regions. There are many possible ways to aggregate from the measurement level to the region level. Ideally, a mixture of statistical and subject-matter considerations determines the final choice of regions to report on. The process of choosing regions of interest is generally too complex to model. Ideally, researchers look at the data in order to choose their regions of interest. How can we do valid inference if researchers looked at data prior to choosing their hypotheses?

In this talk we show how the closed testing method can be used to allow such selective inference by correcting for multiple testing over all (exponentially many) possible regions the researcher might choose. We show that, surprisingly, such unprecedented multiplicity is in fact doable computationally. Even more surprisingly, the resulting method has comparable power compared to competing methods that specify regions of interest as a fixed function of the data. We give a general overview of methods, highlighting the generality of the approach but also going into specific methods e.g. based on the Simes inequality or permutations. We also prove that closed testing is the only possible approach to this problem, showing that methods that are not special cases of closed testing are inadmissible. The practical value of the approach is illustrated with an application on brain imaging data.