

# Using Bayes Factors to Determine Significant Clusters

*Claudio Fuentes*

Department of Statistics, Oregon State University

*Email:* `fuentes@stat.oregonstate.edu`

**Abstract:** The detection and determination of clusters has been an important concern, among researchers from different fields, for a long time. In particular, assessing whether the clusters are statistically significant, is a question that has been asked by a number of experimenters. Although several efforts have been made in this direction, not much has been done with the aim of developing an actual hypothesis test in order to assess the significance of clusters. In this talk, I will present a new methodology that allows the examination of the hypothesis test  $H_0 : K = 1$  vs.  $H_1 : K = k$ , where  $K$  denotes the number of clusters present in a certain population.

This procedure, based on Bayesian tools, permits us to obtain closed form expressions for the posterior probabilities corresponding to the null hypothesis. From here, we calibrate our results by estimating the frequentist null distribution of the posterior probabilities in order to obtain the  $p$ -values associated with the observed posterior probabilities. Finally, I will discuss an application of this method to the analysis of NIR spectroscopy data coming from the maize genetic study that motivated this work..