"Tag-customized Critical Levels for Differential Expression Testing"

Carlos Alberto de Bragança Pereira[1]

[1] Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP)

Abstract

Significance test is a method that indicates de consistency between data and null hypothesis. To take a decision one compares the observed significance level against a critical one. 10%, 5%, and 1% are canonical critical levels used in most statistical analysis in the scientific literature. None can explain why these numbers turned to be a standard. As a consequence of this "nonsense" it is common to hear in the medical academy the following sentence: *increase the sample size to reject the null hypothesis.* Worst is when someone comes to tell us the following: *My master thesis has a small sample but now for the doctorate I will consider a much bigger experiment.* One application of significance test in bioinformatics is the classification of tags in digital expression profiles (DEP), such as SAGE or MPSS. In these experiments it is very important to detect the tags that are differentially expressed. The use of a fixed critical level is inappropriate in the analysis of typical DEP experiment due to the large variation of tag frequencies. In our lecture we want to convince the audience about the need for choosing, for each tag, a critical level that is a function [1] of the total tag frequency, whenever comparing libraries.

References

[1] Varuzza L; Gruber A; Pereira CAB (2008), Significance tests for comparing gene expression profile, arXiv:0806.3274v3 [q-bio.GN]