

**Title:** Multidimensional statistics and applications to study genes

**Author:** Mgr. Peter Bubeník

**Department:** Department of probability and mathematical statistics

**Supervisor:** prof. Lev Klebanov, DrSc., KPMS MFF UK

**Abstract:** Microarray data of gene expressions consist of thousands of genes and just some tens of observations. Moreover, genes are highly correlated between themselves and contain systematic errors. Hence the magnitude of these data does not afford us to estimate their correlation structure. In many statistical problems with microarray data, we have to test some thousands of hypotheses simultaneously. Due to dependence between genes,  $p$ -values of these hypotheses are dependent as well. In this work, we compared convenient multiple testing procedures reasonable for dependent hypotheses. The common manner to make microarray data more uncorrelated and partially eliminate systematic errors is normalizing them. We proposed some new normalizations and studied how different normalizations influence hypotheses testing. Moreover, we compared tests for finding differentially expressed genes or gene sets and identified some interesting properties of some tests such as bias of two-sample Kolmogorov-Smirnov test and interesting behavior of Hotelling's test for dependent components of observations. In the end of this work, we proposed test for testing independence of genes.

**Keywords:** Multiple testing procedures, microarray, gene expressions