Mathematical Statistics
Stockholm University
Research Report **2019:17**,
http://www.math.su.se

Multivariate multiple test procedures

Thorsten Dickhaus[a], André Neumann[a], and Taras Bodnar[b]

[a]Department of Mathematics, University of Bremen, Bremen, Germany

[b]Department of Mathematics, Stockholm University, Stockholm, Sweden

December 2019

Keywords: multiple test; copula; effective numbers of tests; FWER; FDR

1 Introduction and preliminaries

Dependencies among data points are present in virtually all modern statistical applications. This holds especially true for studies with multiple endpoints which are all measured for the same observational units. For example, consider the case of a gene expression study. In that context, expression levels of m genes are measured for n individuals. The goal of the study typically is to detect statistically significant expression differences, either in a two-groups model or in a one-group model under different experimental conditions. Due to biological and technological reasons, the expression levels will typically exhibit strong dependencies, at least for genes which are functionally related; cf. [67]. We will discuss multiple tests which take such dependencies explicitly into account. Such multiple tests are called multivariate multiple tests, because they rely on joint distributions or on approximations thereof.

1.1 Motivation

Example 1. *As a simple motivating example for utilizing a multivariate multiple test, consider the case of $m = 2$ simultaneous tests for Gaussian means. Let $Z = (Z_1, Z_2)^\top$ denote an observable $\mathbb{R}^2$-valued random vector which follows the bivariate normal distribution with an unknown mean vector $\mu = (\mu_1, \mu_2)^\top$, but a known covariance (and correlation) matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where $|\rho| < 1$. Consider the two (one-sided) null hypotheses $H_j = \{\mu_j \leq \mu_j^*\}$, $j = 1, 2$, for a given vector $\mu^* = (\mu_1^*, \mu_2^*)^\top \in \mathbb{R}^2$. The corresponding alternative hypotheses are given by $K_j = \{\mu_j > \mu_j^*\}$, $j = 1, 2$. This is a typical setup for a*