

Název práce: Mnohorozměrná statistika a aplikace na studium genů

Autor: Mgr. Peter Bublín

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí disertační práce: prof. Lev Klebanov, DrSc., KPMS MFF UK

Abstrakt: Microarrayová data genových expresí se skládají z několika tisíců genů a pouze několika desítek pozorování. Navíc, geny jsou mezi sebou silně závislé a data obsahují systematické chyby. Proto nám rozsah těchto dat nedovoluje rozumně odhadnout jejich korelační strukturu. U mnoha statistických problémů s microarrayovými daty musíme současně testovat tisíce hypotéz. Vzhledem k závislosti mezi geny, p -hodnoty těchto hypotéz jsou taky závislé. V této práci porovnáme běžné procedury mnohonásobného testování, které jsou vhodné pro závislé hypotézy. Běžný způsob, jak udělat microarrayová data méně závislá a částečně odstranit systematické chyby, je normalizovat je. Proto bylo navrženo několik nových normalizací a studovali jsme, jak různé normalizace ovlivňují testování hypotéz. Navíc jsme porovnali testy pro nalezení odlišně expresovaných genů nebo genových množin a našli několik zajímavých vlastností testů jako například strannost dvojvýběrového Kolmogorov-Smirnovova testu a zajímavé chování Hotellingova testu pro závislé složky pozorování. Na konci jsme navrhli test pro testování nezávislosti genů.

Klíčová slova: Procedury mnohonásobného testování, microarray, genové exprese