

# Obsah

|                                                                               |           |
|-------------------------------------------------------------------------------|-----------|
| Používané značení                                                             | 8         |
| Úvod                                                                          | 10        |
| Struktura knihy                                                               | 12        |
| <b>1 Teorie chybějících hodnot</b>                                            | <b>14</b> |
| 1.1 Problém chybějících hodnot                                                | 14        |
| 1.2 Vzorce chybějících hodnot                                                 | 18        |
| 1.3 Mechanismy chybějících hodnot                                             | 23        |
| 1.3.1 Technické definice mechanismů                                           | 24        |
| 1.3.2 Mechanismus MCAR – hodnoty chybí zcela náhodně                          | 26        |
| 1.3.2.1 Testování mechanismu MCAR v datech                                    | 27        |
| 1.3.3 Mechanismus MAR – hodnoty chybí náhodně                                 | 30        |
| 1.3.4 Mechanismus NMAR – hodnoty nechybí náhodně                              | 32        |
| 1.3.5 Ignorovatelný a neignorovatelný mechanismus                             | 34        |
| <b>2 Metody práce s chybějícími hodnotami</b>                                 | <b>35</b> |
| 2.1 Metody založené na vynechávání případů                                    | 38        |
| 2.1.1 Analýza kompletních případů                                             | 38        |
| 2.1.2 Analýza dostupných případů                                              | 41        |
| 2.2 Metody založené na nahrazování chybějících hodnot                         | 43        |
| 2.2.1 Nahrazování chybějících hodnot aritmetickým průměrem                    | 43        |
| 2.2.2 Nahrazování chybějících hodnot regresi                                  | 46        |
| 2.2.3 Nahrazování chybějících hodnot stochastickou regresi                    | 48        |
| 2.2.4 Další metody                                                            | 50        |
| 2.3 Metody založené na maximální věrohodnosti                                 | 52        |
| 2.3.1 EM algoritmus                                                           | 54        |
| <b>3 Mnohonásobné imputace</b>                                                | <b>57</b> |
| 3.1 Bayesovská statistika a chybějící hodnoty                                 | 59        |
| 3.2 Věcná analýza s využitím mnohonásobných imputací                          | 63        |
| 3.3 Imputační algoritmy                                                       | 66        |
| 3.3.1 Mnohonásobné imputace založené na mnohorozměrném normálním rozdělení    | 67        |
| 3.3.2 Algoritmus MICE                                                         | 69        |
| 3.4 Slučování výsledků a statistická inference                                | 72        |
| 3.5 Kolik imputací je potřeba provést?                                        | 75        |
| <b>4 Simulace jako nástroj srovnávání metod práce s chybějícími hodnotami</b> | <b>79</b> |
| 4.1 Motivace pro používání simulací při srovnávání metod                      | 79        |
| 4.2 Cíle analytické části                                                     | 80        |

|                                                                        |            |
|------------------------------------------------------------------------|------------|
| 4.3 Postup realizace simulační studie                                  | 83         |
| 4.4 Statistický software a reprodukovatelnost simulací                 | 85         |
| 4.5 Nástroje vyhodnocování výsledků simulací                           | 86         |
| <b>5 První simulační studie: vztah měsíční mzdy a IQ</b>               | <b>90</b>  |
| 5.1 Definování experimentálních podmínek                               | 90         |
| 5.2 Definování populace a generování dat                               | 93         |
| 5.3 Vytváření chybějících hodnot podle jednotlivých mechanismů         | 96         |
| 5.4 Vyhodnocování výsledků                                             | 100        |
| <b>6 Druhá simulační studie: politická znalost a její determinanty</b> | <b>113</b> |
| 6.1 Definování experimentálních podmínek                               | 115        |
| 6.2 Definování populace a generování dat                               | 117        |
| 6.3 Vytváření chybějících hodnot podle jednotlivých mechanismů         | 120        |
| 6.4 Vyhodnocování výsledků                                             | 123        |
| 6.5 Diskuse                                                            | 130        |
| <b>Závěr</b>                                                           | <b>132</b> |
| <b>Summary</b>                                                         | <b>135</b> |
| <b>Použitá literatura</b>                                              | <b>138</b> |
| <b>Rejstřík</b>                                                        | <b>145</b> |